KOLLOQUIUM FREMDSPRACHENUNTERRICHT 51

Daniela Elsner / Britta Viebrock (Hrsg.)

Triangulation in der Fremdsprachenforschung





KOLLOQUIUM FREMDSPRACHENUNTERRICHT 51

Daniela Elsner / Britta Viebrock (Hrsg.)

Triangulation in der Fremdsprachenforschung

Fremdsprachenlernen und Fremdsprachenunterricht sind facettenreiche Forschungsgegenstände, deren Untersuchung ein komplexes Design verlangt. Um ein multidimensionales Bild der ablaufenden Prozesse zu erhalten, werden in der fremdsprachlichen Unterrichtsforschung immer häufiger rekonstruktive und interpretative Verfahren mit standardisierten quantitativen Methoden verbunden. Methoden-, Theorie-, Daten- oder Beobachtertriangulation werden zur Überprüfung von Forschungsergebnissen sowie zur Erweiterung von Erkenntnismöglichkeiten eingesetzt. Die Beiträge in diesem Band zeigen die unterschiedlichen Dimensionen des Triangulationskonzepts, seine theoretischen Grundlagen sowie praktische Anwendungen. Sie sind im Anschluss an die zweite forschungsmethodische Sommerschule der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) entstanden.

Die Herausgeberinnen

Daniela Elsner ist Professorin für Sprachlehrforschung und Didaktik der englischen Sprache an der Goethe-Universität Frankfurt/Main. Ihre Schwerpunkte in Forschung und Lehre sind Mehrsprachigkeit, Multiliteralität, Fremdsprachenlernen in der Grundschule sowie Bilinguale Lehr- und Lernprozesse.

Britta Viebrock ist Professorin für Didaktik der englischen Sprache und Literatur an der Goethe-Universität Frankfurt/Main. Ihre Schwerpunkte in Forschung und Lehre sind CLIL, Multiliteralität, Rekonstruktive Forschung sowie Forschungsethik.

www.peterlang.com

Triangulation in der Fremdsprachenforschung

KOLLOQUIUM FREMDSPRACHENUNTERRICHT

Herausgegeben von Daniela Caspari,
Lars Schmelter, Karin Vogt und Nicola Würffel

BAND 51

Zu Qualitätssicherung und Peer Review der vorliegenden Publikation:

Die Qualität der in dieser Reihe erscheinenden Arbeiten wird vor der Publikation durch alle vier Herausgeber der Reihe geprüft.

Notes on the quality assurance and peer review of this publication:

Prior to publication, the quality of the work published in this series is reviewed by all four editors of the series.

Daniela Elsner / Britta Viebrock (Hrsg.)

Triangulation in der Fremdsprachenforschung



Bibliografische Information Der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über ‹http://dnb.d-nb.de› abrufbar.

Umschlaglogo:
Christoph Baum

ISSN 1437-7829 • ISBN 978-3-631-65517-7 (Print)
E-ISBN 978-3-653-04899-5 (E-PDF) • E-ISBN 978-3-653-97996-1 (EPUB)
E-ISBN 978-3-653-97995-4 (MOBI) • DOI 10.3726/978-3-653-04899-5





Diese Publikation wurde begutachtet.

Peter Lang GmbH
Internationaler Verlag der Wissenschaften
Berlin

www.peterlang.com

Inhalt

Einleitung: Triangulation in der Fremdsprachenforschung

Daniela Elsner / Britta Viebrock

Der Weg zu neuen Modellen und Konzepten im Fremdsprachenunterricht erschließt sich im Optimalfall über eine Verzahnung theoretischer Analysen, systematischer Beobachtungen und empirischer Untersuchungen im Praxisfeld. Dabei stellt sich für alle drei Herangehensweisen die Frage nach geeigneten Methoden, die den jeweiligen Gegenstand in angemessener Art und Weise beleuchten können. Gerade in Deutschland zeichnete sich im Kontext empirischer Herangehensweisen lange Zeit die Tendenz ab, qualitative und quantitative Erhebungsmethoden strikt voneinander zu trennen. In jüngster Zeit lassen sich jedoch immer häufiger Arbeiten finden, welche die Grenzen zwischen den unterschiedlichen Forschungsparadigmen überwinden und im Sinne einer Methoden-, Theorie-, Daten- oder Beobachtertriangulation die Vorteile unterschiedlicher Herangehensweisen durch integrative Verfahren vereinen. Dieser einleitende Beitrag gibt einen kurzen Überblick über traditionelle und jüngere Sichtweisen diesbezüglich sowie eine Vorschau auf die in diesem Band vereinten Beiträge, die sich mit dem Thema Triangulation in der Fremdsprachenforschung theoretisch auseinandersetzen oder diese praktisch anwenden.

1. Die Wahl der Methode: Qualitativ, quantitativ oder beides?

> Methodology is one of the most basic questions for any discipline dealing with human interaction. Language learning and the acquisition of the rules for appropriate speech behavior are no exception. The questions of what constitutes data and how we treat data are fundamental to the entire enterprise of explaining and describing the acquisition of a second language. (Wolfson 1986: 689)

Die Festlegung des methodischen Vorgehens wird auch in der fremdsprachendidaktischen Forschung primär von der Ontologie der jeweiligen Fragestellung bestimmt. So verweist Grotjahn (2003: 493) darauf, dass Forschungsmethoden dem jeweiligen Gegenstand angemessen sein sollen. Der Forschungsgegenstand „Fremdsprachenunterricht“ weist dabei – in Abgrenzung zu anderen Forschungsbereichen – eine Reihe von spezifischen Aspekten auf, welche vom Forscher[1]

1 In den Beiträgen dieses Bandes werden maskuline und feminine Formen in unterschiedlicher Weise verwendet. Unter der Voraussetzung, dass jeweils beide

berücksichtigt werden müssen. So zeigen die fremdsprachlichen Interaktionen im Unterricht ganz besondere Eigenschaften, wie z.B. ein hoher Sprechanteil der Lehrkräfte, häufige Fehlerkorrektur, *Code-Switching* etc. auf, die in dieser Art in anderen Forschungsfeldern nicht wiederzufinden sind und deren wissenschaftliche Analyse eine sorgsame Auswahl einer oder mehrerer Forschungsmethoden bedarf.

In der Sozialforschung konkurrieren eine Vielzahl von miteinander nicht oder lediglich partiell kompatiblen Theorien zur Methodologie, welche auch in die fremdsprachliche Unterrichtsforschung Eingang finden. Die Unterscheidung von qualitativer und quantitativer bzw. rekonstruktiver und Hypothesen testender Forschung (Bonnet 2009) scheint dabei eine übergeordnete Rolle zu spielen: Während das qualitative Forschungsparadigma eine holistische Betrachtung der Realität zugrunde legt, indem alle sich möglicherweise auf den Forschungsgegenstand auswirkenden Einflussfaktoren mit einbezogen werden, wird in quantitativen Ansätzen versucht, die Anzahl dieser Faktoren möglichst gering zu halten und den Untersuchungsgegenstand weitgehend zu kontrollieren. Die schriftliche bzw. mündliche Datenerhebung findet dabei mittels ausgewählter Instrumente in einmaligen „Überprüfungssituationen“ statt. Quantitative Forschungsdesigns überprüfen vorab formulierte Hypothesen mittels einer großen Anzahl an Probanden. Qualitative Untersuchungen sind dagegen Hypothesen generierend, explorativ, rekonstruktiv und häufig in Longitudinalstudien angelegt. In der quantitativen Forschung wird der Unabhängigkeit des Beobachters vom Forschungsgegenstand ein zentraler Stellenwert eingeräumt. Qualitative Forschung greift dagegen auf die methodisch kontrollierte, subjektive Wahrnehmung des Forschers als Bestandteil der Erkenntnisgewinnung zurück (vgl. z.B. Flick et al. 2003: 24f.).

Jenseits der Debatten, in denen sich beide Forschungsrichtungen wechselseitig die wissenschaftliche Legitimation absprechen, muss vom Forscher zunächst überlegt werden, für welche Fragestellung und für welchen Forschungsgegenstand qualitative oder quantitative Forschung jeweils angemessener erscheint. Quantitative Methoden messen zählbare Eigenschaften. Die häufigsten Methoden der quantitativen Datenerhebung sind die Befragung, die Beobachtung, das Experiment, physiologische Messungen sowie der Test (vgl. Bortz/Döring 2002). Quantitative Verfahren sind für ihre vergleichend-statistische Auswertung auf ein hohes Maß an Standardisierung der Datenerhebung angewiesen. Der Vorteil

Geschlechter gemeint sind, haben wir uns entschieden, diese Vielfalt bestehen zu lassen. Wir fühlen uns den Leitprinzipien des *gender mainstreaming* verpflichtet, sehen sie jedoch nicht notwendigerweise dadurch eingelöst, dass jeder maskulinen Form noch ein „/-innen“ beigefügt wird.

quantitativer Methoden liegt darin, dass sich die Messergebnisse mit algebraischen (Addition, Multiplikation) beziehungsweise den darauf basierenden stochastischen Methoden (Mittelwert, Varianz, etc.) weiterverarbeiten lassen. So ist es möglich, statistische Tests durchzuführen und Hypothesen zu prüfen sowie deren Signifikanz zu berechnen. Explorativ lassen sich Daten zudem mittels Faktorenanalyse oder Clusteranalyse auswerten (vgl. Elsner/Wildemann 2010).

Lange Zeit stand die qualitative Sozialforschung den Erhebungsmethoden der quantitativen Forschung sehr kritisch gegenüber. Insbesondere die Künstlichkeit der Erhebungssituation, eine mangelnde Offenheit und die Nichtberücksichtigung sozialer Phänomene wurden als Defizite standardisierter Verfahren angesehen. Qualitative Forschung versucht hingegen, Soziales tiefgründiger zu erforschen und nicht direkt sichtbare Sinnstrukturen herauszuarbeiten. Die damit einhergehenden Verfahren lassen sich stärker dem Einzelfall anpassen und sind damit in Bezug auf die Erhebung individueller Verläufe wesentlich flexibler. Zu den häufigsten qualitativen Methoden der Datenerhebung zählen die schriftliche und mündliche Befragung sowie die qualitative Beobachtung. Qualitative Verfahren sind immer dort zu empfehlen, wo es um die Erschließung eines bisher wenig erforschten Wirklichkeitsbereichs mithilfe von sensibilisierenden Konzepten geht (vgl. Flick et al. 2003: 25).

Obwohl insbesondere in der deutschen Forschungslandschaft lange Zeit eine starke Tendenz vorherrschte, qualitative und quantitative Methoden zwei unterschiedlichen Paradigmen zuzuordnen (vgl. z.B. Kelle/Erzberger 2003: 299), scheint sich die Grenze zwischen diesen beiden Methoden mittlerweile aufzulösen. So versuchen zahlreiche Arbeiten methodologische und methodische Grundlagen für eine Integration beider Ansätze zu entwickeln. Sie verfolgen dabei im Wesentlichen zwei Zielvorstellungen: (a) eine Perspektivenerweiterung oder (b) eine Ergebnisvalidierung. Während das Ziel einer Perspektivenerweiterung grundsätzlich mit der Forderung nach der Gegenstandsangemessenheit der Methode vereinbar ist, lassen sich bezüglich der Methodenintegration als Validierungsstrategie auch kritische Positionen finden:

> Der Gegenstand nimmt […] unterschiedliche Gestalt an, je nachdem ob er in Form eines Beobachtungsprotokolls, eines Transkriptes oder eines aufgezeichneten Interviews vorliegt, was darauf hindeutet, dass eine durchgängige Gegenstandskonstruktion durch die jeweilige Spezifik der Methode entsteht. (Helsper et al. 2001: 257)

Diese Sichtweise beinhalte zugleich eine fundamentale Kritik an Verfahren, bei denen Ergebnisse, die mithilfe *einer* Forschungsmethode erzeugt werden, durch solche, die mithilfe einer *anderen* hervorgebracht werden, validiert werden. Im strengen Sinne werden hierbei niemals die gleichen Gestalten des Gegenstands zueinander in Beziehung gesetzt (vgl. Viebrock 2007: 34f.). Teilt man diese Kritik,

ist die Zielvorstellung der Ergebnisvalidierung ausgeschlossen. Der Wert eines integrativen Verfahrens zur Perspektivenerweiterung bleibt von der Kritik allerdings unbenommen.

2. Beispiele methodischer Triangulation im Kontext fremdsprachlicher Unterrichtsforschung

Die kombinierte Verwendung zweier oder mehrerer Methoden wird im Kontext der empirischen Sozial- und Geisteswissenschaften als *integrativer Ansatz* (Seipel/Rieker 2003), *mixed methods approach* (Jick 1979) oder *Triangulation* (Flick 2011) bezeichnet. In Anlehnung an Denzin (1970) und andere lassen sich folgende vier Formen der Triangulation unterscheiden:

- Datentriangulation: Kombination und Nutzung mehrerer Datenquellen, die zu unterschiedlichen Zeiten und Orten sowie an verschiedenen Personen erhoben werden.
- Beobachtertriangulation: Datenerhebung und -analyse verschiedene Forscher bzw. Beobachter, um möglichen subjektiven Einflüssen in der Interpretation entgegenzuwirken.
- Theorie-Triangulation: Anwendung unterschiedlicher Theorien auf denselben Forschungsgegenstand.
- Methodentriangulation: Diese kann sowohl in Form der Verwendung verschiedener Skalierungsverfahren innerhalb einer Methode bzw. eines Messinstruments (*within-method*) erfolgen, als auch im Einsatz verschiedener Methoden bei der Datengewinnung zu einem Untersuchungsgegenstand bestehen (*between-method*).

Das generelle Anliegen dieser Forschungsvarianten besteht darin, die besonderen Stärken und Potenziale der gewählten Methoden bzw. Daten in gegenseitiger Ergänzung nutzbar zu machen und die Schwächen zu kompensieren: „Dabei verlagert sich insgesamt der Fokus von der ursprünglichen Orientierung an der Überprüfung von Ergebnissen […] zugunsten einer stärkeren Betonung der systematischen Erweiterung der Erkenntnismöglichkeiten" (Flick 2011: 26).

Das hierzu, nicht nur im Kontext der Unterrichtsforschung, häufig eingesetzte, klassische Phasenmodell einer sequenziellen Verbindung von qualitativen und quantitativen Methoden ist üblicherweise mit der Ambition verbunden, in einem ersten Schritt das Potenzial qualitativer Verfahren für die Exploration des Untersuchungsfelds und die Entwicklung von Instrumenten bzw. Hypothesen zu nutzen, um dann in einem zweiten Schritt mithilfe quantitativer Methoden präzise Messungen durchzuführen und auf der Grundlage einer umfangreicheren

Datenbasis geeignete statistische Analysen vorzunehmen (vgl. Barton/Lazarsfeld 1979, Kelle/Erzberger 2003).

In der fremdsprachlichen Unterrichtsforschung werden immer häufiger interpretative qualitative Verfahren mit standardisierten quantitativen Methoden zu gemeinsamen Untersuchungsdesigns verbunden in der Hoffnung, ein multidimensionales Bild der im Unterricht ablaufenden Prozesse zu erhalten und diese langfristig verändern zu können. Forschung geschieht somit nicht zum Selbstzweck, sondern findet stets zweckgerichtet statt. So versuchte die 2006 abgeschlossene DESI-Studie (Deutsches Institut für Internationale Pädagogische Forschung – DIPF 2006) ebenso wie die im Jahr 2009 veröffentlichte EVENING-Studie (Engel et al. 2009) oder die vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) bundesweit durchgeführte Ländervergleichsstudie (vgl. Köller et al. 2010), mit variierenden Forschungsmethoden unterrichtliche Bedingungsfaktoren für sprachliche und kulturelle Lernfortschritte zu identifizieren. Die Auswertung erfolgte sowohl auf der Basis der im Sinne der Methodentriangulation erhobenen Ergebnisse von standardisierten Leistungstests und/oder Unterrichtsbeobachtungen in Form von Videografien als auch auf der Grundlage von Befragungen aller am Unterricht Beteiligten. Darüber hinaus wurden die bundesweiten Erhebungen sprachlicher Leistungen im Deutschen (als Muttersprache) und im Englischen (als Fremdsprache) (DESI und Ländervergleich) ebenso wie die bundeslandinterne EVENING-Evaluation interdisziplinär angelegt (Fachdidaktiker, Soziologen, Pädagogen, Psychologen); man setzte somit nicht nur auf theoretische und methodische Triangulation, sondern auch auf Forscher- bzw. Beobachtertriangulation. Langfristiges Ziel dieser kaleidoskopartigen Bestandsaufnahmen ist es, die Lernbedingungen und somit die Qualität des fremdsprachlichen Unterrichts auf allen Ebenen zu verbessern.

3. Zum Aufbau und zu den Beiträgen dieses Bandes

Die Beiträge in diesem Band bilden die unterschiedlichen Dimensionen des Triangulationskonzepts aus der Perspektive der Fremdsprachenforschung ab. Sie sind im Anschluss an die zweite forschungsmethodische Sommerschule der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) ausgearbeitet worden. Die fundierte methodische Ausbildung von Nachwuchsforscherinnen und -forschern gehört zu den zentralen Zielsetzungen der DGFF, die mit der Sommerschule zu diesem Zweck ein spezifisches Förderprogramm aufgelegt hat. Den Überblicksbeiträgen etablierter Wissenschaftlerinnen und Wissenschaftler haben wir jeweils einen Anwendungsbeitrag zur Seite gestellt, in dem eine Nachwuchsforscherin bzw. ein Nachwuchsforscher ihr bzw. sein laufendes

Forschungsprojekt, d.h. Fragen des Designs, methodische Entscheidungen und erste Ergebnisse diskutiert.

In Anlehnung an die weit verbreitete Reklame eines bekannten Einrichtungshauses, das mit seinen Möbeln nicht nur zweckmäßige und alltagstaugliche Gegenstände, sondern ein umfassendes, modernes Lebensgefühl verkaufen will, stellt Julia Settinieri in ihrem Beitrag „Forschst Du noch, oder triangulierst Du schon?" einleitend Überlegungen an, ob Triangulation in der Fremdsprachenforschung bereits unverzichtbar ist, wenigstens aber zum guten Ton gehört. In ihrem Überblicksbeitrag zeichnet sie die Genese des Konzepts und seiner unterschiedlichen Ausprägungen nach. Sie zeigt anhand spezifischer Beispiele aus der Fremdsprachenforschung auf, in welcher Weise sich Triangulationsstrategien innerhalb eines Forschungsparadigmas oder übergreifend sinnvoll einsetzen lassen, berücksichtigt allerdings auch die Frage, inwieweit es sich dabei um eine kritisch zu reflektierende „Modeerscheinung" handelt.

Sonja Brunsmeier diskutiert in ihrem Praxisbeitrag ein auf Triangulation basierendes Forschungsdesign zur Frage, wie sich die Entwicklung der in den Bildungsstandards verankerten Interkulturellen Kommunikativen Kompetenz (in Anlehnung an Byram 1997) im Englischunterricht der Grundschule untersuchen lässt. Während das Konzept für die weiterführenden Schulen bereits ausgearbeitet ist und Eingang in Forschungsarbeiten gefunden hat (z.B. im Rahmen der DESI-Studie durch Göbel 2007), sind für das frühe Fremdsprachenlernen noch konzeptionelle und methodische Überlegungen zur Anbahnung von IKK zu leisten, entsprechende Aufgaben zu entwickeln und einer empirischen Überprüfung zu unterziehen. Mithilfe eines komplexen Designs versucht Sonja Brunsmeier, die Perspektiven der unterschiedlichen Akteure zu erhellen.

Die videogestützte Unterrichtsforschung steht im Zentrum des Beitrags von Johannes Appel und Udo Rauin, die sich die neueren technischen Möglichkeiten zur Weiterentwicklung des methodischen Repertoires zur Erforschung von Lehr- und Lernprozessen zunutze machen. Neben einem Überblick über die Grundsätze videographischer Verfahren in der Unterrichtsforschung werden Datenbeispiele aus einem Forschungsprojekt zur Schülerbeteiligung und zum Engagement im Englischunterricht vorgestellt, um die Analyseschritte (Rating bzw. Kodierung) zu unterschiedlichen Aspekten (Qualität des Unterrichts, Sprechanteile, unterrichtliche Sozialformen, Beteiligungsverhalten) transparent zu machen. Die Autoren zeigen auf, dass sich Videodaten insbesondere zur Datentriangulation und zur Methodentriangulation (*within-method*) eignen.

Im zugehörigen Praxisbeitrag zeigt Heidi Seifert sehr überzeugend, wie sich „Videografie als Instrument zur Erforschung von Interaktionsprozessen im Elementarbereich" einsetzen lässt und welche spezifischen Planungsschritte dafür

nötig sind. In einer bilingualen (deutsch-englischen) Kindertagesstätte untersucht die Verfasserin die sprachlichen Besonderheiten der Erzieherin-Kind-Interaktionen und diskutiert Aspekte der Vorbereitung und Durchführung der Datenerhebung. Im Mittelpunkt der Ausführungen stehen die Entwicklung und Reflexion eines geeigneten Aufnahmekonzepts ebenso wie datenrechtliche Überlegungen. Erste Ausblicke auf den Analyseprozess werden darüber hinaus gegeben.

Wie sich die Strategie der Triangulation mithilfe qualitativer Forschungssoftware verfolgen und unterstützen lässt, ist Thema des Grundsatzbeitrags von Claus Stefer. Anhand der Software MAXQDA 11 erörtert der Verfasser mit großem Praxisbezug das Potenzial moderner Analyseprogramme, große Datenmengen systematisch zu verwalten und unterschiedliche Forschungsmethoden zu integrieren. Nach der Darstellung einiger Grundsatzüberlegungen zur Triangulation erläutert der Verfasser einzelne Verfahrensschritte (Codierungen, Verteilungshäufigkeiten von Variablen, Joint Displays, Intercoder-Übereinstimmung) und ihre Umsetzung in der Software. Zahlreiche Screenshots illustrieren die Arbeitsweise des Programms.

Der zugehörige Praxisbeitrag „'Are you good at reading?' Zur Diagnosekompetenz von Englischlehrkräften" von Laura Armbrust verdeutlicht, wie die qualitative Inhaltsanalyse sowie die Auswertungssoftware in einem fremdsprachendidaktischen Forschungsvorhaben zur Anwendung kommen. Insbesondere die wenig zufriedenstellenden Leistungen deutscher Schülerinnen und Schüler in den großen Bildungsstudien haben die diagnostischen Kompetenzen der Lehrkräfte auch mit Blick auf die rezeptiven Fertigkeiten in den Mittelpunkt des Interesses gerückt. Anhand der Daten einer fokussierten Interviewstudie, die von der Verfasserin als *within-method* Triangulation konzipiert wird, werden systematisch die Verfahrensschritte des Auswertungsprozesses aufgezeigt.

Der Grundsatzbeitrag „Analyse qualitativer und quantitativer Befragungsdaten mit SPSS" von Astrid Jurecka nähert sich dem Thema Triangulation aus der Perspektive zweier prototypischer Methoden zur Befragung: dem Interview und dem Fragebogen. Die erstgenannte wird üblicherweise dem qualitativen Paradigma zugeordnet, die letztgenannte dem quantitativen. Neben einer Darstellung der klassischen Verwendungsformen beider Methoden richtet die Verfasserin ihren Blick insbesondere auf die Möglichkeiten der Quantifizierung qualitativer Daten. Anhand von Beispielen aus der Fremdsprachenforschung zeigt sie zudem auf, welche Berechnungen mit diesen Daten mittels der Software SPSS möglich sind.

Der zugehörige Praxisbeitrag von Sara Dallinger und Kathrin Jonkmann stellt eine Längsschnittstudie zum bilingualen (deutsch-englischen) Geschichtsunterricht vor: „Competencies and Motivation in the Bilingual History Classroom (COMBIH)", in der zu zwei Messzeitpunkten quantitativ orientierte

Erhebungsinstrumente (Kompetenztests, Fragebögen) eingesetzt werden. Im Zentrum der Überlegungen der Verfasserinnen steht die Frage, welche triangulativen Optionen sich aus ihren Datensätzen ergeben. Sie zeigen, wie sich mithilfe von Forscher- und insbesondere Datentriangulation sowohl die Strategie der Validierung als auch die Strategie der Perspektivenerweiterung verfolgen lässt.

Der abschließende Beitrag von Karin Aguado zeigt die Grenzen von Triangulationsstrategien auf und weist auf zahlreiche kritische Punkte hin, die es bei der Planung und Durchführung von triangulativ angelegten Forschungsvorhaben abzuwägen gilt. Im Zentrum der Argumentation dieses Beitrags steht die Annahme, dass das methodische Design eines jeden Forschungsprojekts zunächst aus der Gegenstandsangemessenheit der gewählten Methode(n) zu begründen ist. Diese wird nicht *per defintionem* durch den Ansatz der Triangulation erhöht, welcher auch nach Ansicht von Karin Aguado gelegentlich Gefahr läuft, zu einer „Modeerscheinung“ zu degenerieren. Allerdings liegen in den Anforderungen, die eine fundierte Triangulationsstrategie mit sich bringt, gerade für die Fremdsprachenforschung besondere Potenziale, kooperative Vorhaben umzusetzen und kollektive Datenbanken anzulegen, welche von einer größeren Anzahl von Forscherinnen und Forschern systematisch genutzt werden können und somit einen deutlich sichtbareren Beitrag zur Entwicklung der Forschungsleistungen der Disziplin liefern.

Mit dem abschließenden Beitrag, der gezielt auch eine kritische Sicht der zuvor recht positiv dargestellten Potenziale von Triangulation aufgreift, wird der Band in seinen unterschiedlichen Perspektiven abgerundet: Neben den Möglichkeiten, die mit der Triangulation verbunden sind, den Ausprägungen, mit denen sie umgesetzt werden kann, den Ansprüchen, denen sie genügen muss, den praktischen Entscheidungen, die zu erwägen sind, werden auch Schwierigkeiten und Einschränkungen nicht verschwiegen. Mit dieser Ausrichtung verbinden wir die Hoffnung, Wissenschaftlerinnen und Wissenschaftlern sowie insbesondere Nachwuchskräften einen differenzierten Reflexionsrahmen zur Verfügung zu stellen, der für die Planung neuer Forschungsdesigns zahlreiche Argumente zur Verfügung stellt.

Literatur

Barton, Alan H. / Lazarsfeld, Paul F. (1979). Einige Funktionen von qualitativer Analyse in der Sozialforschung. In: Hopf, Christel / Weingarten, Elmar (Hrsg.). *Qualitative Sozialforschung*. Stuttgart: Klett-Cotta, 41-89.

Bonnet, Andreas (2009). Rekonstruktive Methodologie in der empirischen Fremdsprachenforschung: Drei Antworten auf einen gängigen Einwand. In: Lütge, Christiane / Kollenrott, Anne / Ziegenmeyer, Birgit / Fellmann, Gabriela

(Hrsg.). *Empirische Fremdsprachenforschung – Konzepte und Perspektiven*. Frankfurt/Main: Lang, 23-37.

Bortz, Jürgen / Döring, Nicola (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 3. Aufl. Göttingen: Hogrefe.

Byram, Michael (1997). *Teaching and Assessing Intercultural Competence*. Clevedon: Multilingual Matters.

Creswell, John (2003). Research Design. Qualitative, Quantitative and Mixed Methods Approaches. 2. Aufl. Thousands Oaks: Sage.

Denzin, Norman K. (1970). *The Research Act*. Chicago: Aldine.

Deutsches Institut für Internationale Pädagogische Forschung (DIPF) (2006). Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englische-Schülerleistungen-International (DESI). Frankfurt/Main. Online: www.dipf.de/desi/DESI_Zentrale_Befunde.pdf, 61 S. [10.3.2014].

Elsner, Daniela / Wildemann, Anja (2010). Empirie und Fachdidaktik? Überlegungen zur Anwendung quantitativer und qualitativer Forschungsmethoden in der Masterarbeit. In: Elsner, Daniela / Wildemann, Anja (Hrsg.). *Deutsch- und Englischunterricht empirisch betrachtet*. Aachen: Shaker, 5-23.

Engel, Gaby / Groot-Wilken, Bernd / Thürmann, Eike (2009). *Englisch in der Primarstufe –Chancen und Herausforderungen*. Evaluation und Erfahrungen aus der Praxis. Berlin.

Flick, Uwe. (2011). *Triangulation. Eine Einführung*. 3. Aufl Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick, Uwe / Kardorff von Ernst / Steinke, Ines (2003). Was ist qualitative Forschung? Einleitung und Überblick. In: Flick, Uwe / Kardorff von Ernst / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch*. 2. Aufl. Reinbek bei Hamburg: Rowohlt, 13-29.

Göbel, Kerstin (2007). *Qualität im interkulturellen Englischunterricht – eine Videostudie*. Münster: Waxmann.

Grotjahn, Rüdiger (2003). Empirische Forschungsmethoden. Überblick. In: Bausch, Karl-Richard / Christ, Herbert / Krumm, Jürgen (Hrsg.). *Handbuch Fremdsprachenunterricht*. 3. Aufl. Tübingen: Francke, 457-459.

Helsper, Werner / Herwartz-Emden, Leonie / Terhart, Ewald (2001). Qualität qualitativer Forschung in der Erziehungswissenschaft. Ein Tagungsbericht. *Zeitschrift für Pädagogik* 47/2, 251-269.

Jick, Todd D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly* 24/4, 602-611

Kelle, Udo / Erzberger, Christian (2003). Quantitative und Qualitative Methoden – kein Gegensatz. In: Flick, Uwe / Kardorff von Ernst / Steinke, Ines (Hrsg.): *Qualitative Forschung. Ein Handbuch*. 2. Aufl. Reinbek bei Hamburg: Rowohlt, 299-309.

Köller, Olaf / Knigge, Michel / Tesch, Bernd (2010). *Sprachliche Kompetenzen im Ländervergleich.* Münster: Waxmann.

Kuckartz, Udo (2008). Realizing Mixed-Methods Approaches with MAXQDA. Online: http://www.maxqda.de/download/MixMethMAXQDA-Nov01-2010.pdf, 22 S., . [10.3.2014].

Tashakkori, Abbas / Teddlie, Charles (Hrsg.) (2003). *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks: Sage.

Seipel, Christian / Rieker, Peter (2003). Integrative Sozialforschung. Konzepte und Methoden der qualitativen und quantitativen empirischen Forschung. Weinheim/München: Juventa.

Viebrock, Britta (2007). *Bilingualer Erdkundeunterricht. Subjektive didaktische Theorien von Lehrerinnen und Lehrern*. Frankfurt/Main: Lang.

Wolfson, Nessa (1986). Research methodology and the question of validity. *TESOL Quarterly* 20/4, 689-699.

Forschst Du noch, oder triangulierst Du schon?

Julia Settinieri

Die Forschungsstrategie der Triangulation wird aktuell sehr facettenreich diskutiert. Der vorliegende Beitrag zeichnet zentrale Diskussionslinien nach, wobei schwerpunktmäßig auf mögliche Zielsetzungen und Formen, Forschungsparadigmen übergreifende Triangulation sowie ausgewählte forschungspraktische Probleme eingegangen wird. Argumentiert wird, dass der nicht zu unterschätzende Mehraufwand, der durch den Einsatz triangulativer Strategien zwangsläufig entsteht, sich lohnt, wenn Gegenstandsangemessenheit und Messgüte jedes einzelnen Datensatzes gewährleistet sind und eine explizit begründete sowie möglichst tiefgreifende Integration der einzelnen Datensätze gelingt.

1. Einleitung

Triangulation ist in den zwei vergangenen Jahrzehnten in der Sozialforschung und in den letzten Jahren nun auch in der Sprachlehr- und -lernforschung als eine Art Modeerscheinung stark in den Fokus der Fachdiskussion getreten (vgl. bereits Aguado/Riemer 2001: 246). Dabei wirkt das Konzept einerseits schillernd und gleichzeitig seltsam diffus und wird wohl häufiger diskutiert als tatsächlich angewandt, worauf auch die ILMES-Definition des Begriffs *Triangulation* hindeutet:

> In der empirischen Sozialforschung: Die Betrachtung eines Gegenstandes aus (nimmt man den Begriff wörtlich) zwei oder (im übertragenen Sinn) mehreren Perspektiven, Blickrichtungen, Standpunkten. Vor allem im Bereich der qualitativen Forschung wird dieses Verfahren *immer wieder propagiert und gelegentlich sogar eingesetzt.* (ILMES o. J., Hervorhebung J. S.).

Ziel dieses Aufsatzes ist es daher, einen Überblick über Auffassungen und Formen von Triangulation zu geben, um davon ausgehend zu diskutieren, wann der Einsatz triangulativer Forschungsformen sinnvoll sein kann und wann nicht. Der Beitrag gliedert sich in sieben Abschnitte: Im Anschluss an Einleitung und Klärung grundlegender Begriffe werden unterschiedliche Formen von Triangulation, auch Paradigmen übergreifend, diskutiert, um eine Gesamttypologie des Triangulationsbegriffs abzuleiten. Nach der Diskussion einiger mit Triangulation verbundener Praxisprobleme wird abschließend ein die vorangehenden Überlegungen zusammenführendes Fazit gezogen.

2. Definition, Begriffsgeschichte und Zielsetzungen

Ursprünglich wurde die Triangulationsmetapher aus den Fachgebieten der Navigation und der Landvermessung entlehnt (Kelle/Erzberger 2004: 302), wie Brown und Rodgers anschaulich erläutern:

> In surveying and navigation, one determines the position of an object by measuring the angles of observation to this object from two points of already known positions. In the social sciences, triangulation refers to the attempt to understand some aspect of human behavior by studying it from more than one standpoint […]. (Brown/Rodgers 2002: 243)

Diese Grundidee kann jedoch nicht in ihrer Gänze auf die Sozialwissenschaften übertragen werden. Während beispielsweise in der Navigation das Ziel von Triangulation darin besteht, den Standort eines Schiffs im Koordinatensystem durch das Anpeilen des Schiffs von mehreren bekannten Positionen aus zu bestätigen, also quasi mehrfach exakt dasselbe zu messen, muss Triangulation in den Sozialwissenschaften nicht unbedingt auf dasselbe zielen, sondern kann auch gerade auf unterschiedliche Aspekte eines Gegenstands abheben.

> Die „Berechnung der Position eines Ortes durch die Messung von unterschiedlichen Punkten aus" kann hier bedeuten,
>
> 1. dass mit verschiedenen Methoden dasselbe soziale Phänomen erfasst wird oder
> 2. dass hiermit *unterschiedliche Aspekte desselben Phänomens* oder sogar *unterschiedliche Phänomene* erfasst werden, deren Abbildungen sich allenfalls zu einem einheitlichen Bild ergänzen.
>
> Diese Unterscheidung ist keineswegs ein sprachlicher Kunstgriff; denn nur dann, wenn sich verschiedene Methoden auf denselben Gegenstand beziehen, können sie zur wechselseitigen Validierung ihrer Ergebnisse eingesetzt werden. Wenn dagegen verschiedene Methoden verschiedene Aspekte desselben Gegenstandes oder auch unterschiedliche Gegenstände erfassen, so sind unterschiedliche Ergebnisse natürlich zu erwarten, ohne dass dies den Schluss auf deren fehlende Validität erlaubt. (Kelle/Erzberger 2004: 303, Hervorhebung im Original)

In den Sozialwissenschaften lassen sich somit im Rahmen triangulativer Designs zwei ganz unterschiedliche Zielsetzungen verfolgen: Einerseits kann nach Gleichem gesucht werden, um die Validität von Forschungsergebnissen zu bestätigen; andererseits kann aber auch nach ganz unterschiedlichen, einander ergänzenden Informationen gesucht werden, um ein vollständigeres Gesamtbild eines Forschungsgegenstands zu zeichnen (vgl. auch Dörnyei 2007: 164-166). Beide Zielsetzungen sind legitim, haben sich forschungsgeschichtlich jedoch nacheinander entwickelt (vgl. Johnson/Onwuegbuzie 2004, Johnson et al. 2007, Aguado

2014, für einen forschungshistorischen Einblick), wobei sich der Schwerpunkt triangulativer Forschung verschoben hat. Ursprünglich im Rahmen quantitativer Forschung entwickelt, sollte das Triangulationskonzept dazu beitragen, die Beeinflussung von Forschungsergebnissen durch das jeweilige Messinstrument durch den Einsatz mehrerer Messinstrumente zu relativieren und so die Validität der Ergebnisse zu steigern (Reaktivität von Messungen; vgl. genauer z.B. Lamnek 2005: 276-278). Auch Denzin (1970), der für die Verbindung quantitativer und qualitativer Methoden plädiert, tut dies ursprünglich mit dem Ziel ihrer gegenseitigen Validierung. Diskutiert werden im Vorfeld dieser Grundlagenpublikation zur Triangulation bereits der Ansatz der *multitrait-multimethod matrix* (Campbell/Fiske 1959) und der Einsatz sog. non-reaktiver Verfahren, letzteres auch schon unter dem Begriff der *Triangulation* (Webb et al. 1966). Erst ab den 1980er Jahren wird stärker betont, dass Triangulation weniger zur Validierung als zur gegenseitigen Ergänzung von Forschungsansätzen sinnvoll sein könnte (Fielding/Fielding 1986), und zwar unter Rückgriff auf das gleiche Argument: Wenn die Methode den Gegenstand konstituiert, ist es schlicht nicht zu erwarten, dass mit unterschiedlichen Methoden dasselbe gemessen werden kann (vgl. auch die Diskussion in Abschnitt 4). Auch Denzin selbst betont dies in neueren Arbeiten:

> [T]he use of multiple methods, or triangulation, reflects an attempt to secure an in-depth understanding of the phenomenon in question. Objective reality can never be captured. Triangulation is not a tool or a strategy of validation, but an alternative to validation […]. The combination of multiple methods, empirical materials, perspectives and observers in a single study is best understood, then, as a strategy that adds rigor, breadth, and depth to any investigation […]. (Denzin/Lincoln 1994: 2)

Validierung als Zielsetzung von Triangulationsstrategien tritt somit stark in den Hintergrund, bleibt aber in einigen Zusammenhängen, z. B. im Rahmen der Messung von Interrater-Reliabilität, durchaus noch aktuell. Auch für unsere Fächergruppe, deren Forschungsgegenstand sich durch eine besondere Faktorenkomplexion auszeichnet, scheint insbesondere der Aspekt der Erkenntniserweiterung attraktiv. So zeigt Riemer (2011) am Beispiel der Fragestellung *Wann und wie sollen in der mündlichen Sprachproduktion auftretende Grammatikfehler im Fremdsprachenunterricht korrigiert werden?*, dass zu ihrer Beantwortung sowohl lernersprachliche, als auch spracherwerbsbezogene, soziokulturelle und affektive, didaktisch-methodische sowie sprachpolitische Dimensionen in die Überlegungen einbezogen werden müssen, was eine triangulative Herangehensweise nahezu obligatorisch erscheinen lässt. Zusammenfassend kann festgehalten werden:

> Zwei Lesarten der Triangulationsmetapher liegen also vor: Triangulation als kumulative Validierung von Forschungsergebnissen und Triangulation als Ergänzung von Perspektiven, die eine umfassendere Erfassung, Beschreibung und Erklärung eines Gegenstandsbereichs ermöglichen, wobei in der neueren Literatur der Aspekt der Komplementarität, das heißt der Ergänzung von Perspektiven gegenüber dem Aspekt der Validierung hervorgehoben wird: […]. (Kelle/Erzberger 2004: 303-304)

Mit den genannten Unterschieden in der Zielsetzung verbunden sind auch mögliche bzw. jeweils erwünschte Ergebnisse von Triangulation. Theoretisch möglich sind die Feststellung von Konvergenz/Kongruenz, Divergenz und Komplementarität in den erhobenen Daten (Kelle/Erzberger 2004: 304-307, Lamnek 2005: 285-289, Flick 2011b: 49). Wenn Validierung Ziel der Triangulation ist, weist Divergenz auf methodische Probleme hin, Komplementarität auf einen Denkfehler im Design und nur Konvergenz wäre wünschenswert. Sollen beispielsweise qualitative Pretests in Form von Lautdenkprotokollen prüfen, inwiefern Fragebogenitems tatsächlich messen, was sie messen sollen (Kelle/Erzberger 2004: 305), wäre Konvergenz zwischen den Lautdenkdaten und den Konstruktionsintentionen des Forschers[1] wünschenswert, Komplementarität oder gar Divergenz würden eine Überarbeitung der Items notwendig machen. Zu bedenken ist allerdings, dass auch Konvergenz lediglich aufgrund wahrscheinlichkeitstheoretischer Überlegungen das in Forschungsergebnisse gesetzte Vertrauen erhöht, wobei grundsätzlich möglich bleibt, dass auch übereinstimmende Befunde falsch sein können (Lamnek 2005: 285-286).

Ist hingegen Erkenntniserweiterung Ziel der Triangulation, sind vor allem Komplementarität, aber auch Divergenz theoriebefruchtend, Konvergenz hingegen ist weniger interessant. So können qualitative Untersuchungen beispielsweise dabei helfen, „Lücken <variablensoziologischer> Erklärungen, bei denen statistische Zusammenhänge durch zusätzliche *Ex-post*-Annahmen erklärt werden, zu schließen“ (Kelle/Erzberger 2004: 306, Hervorhebung im Original). Vorstellbar wäre, quantitativ festgestellte Unterschiede zwischen L1-Gruppen (z. B. Leistungsunterschiede in einem Sprachtest) im Rahmen qualitativer Folgeuntersuchungen auf ihre denkbaren Ursachen (z. B. Motivationsunterschiede o. Ä.) hin zu untersuchen. Divergenzen, die nicht durch Methodenfehler zu erklären sind, können darüber hinaus zur Modifikation

---

1 In den Beiträgen dieses Bandes werden maskuline und feminine Formen in unterschiedlicher Weise verwendet. Unter der Voraussetzung, dass jeweils beide Geschlechter gemeint sind, haben die Herausgeberinnen entschieden, diese Vielfalt bestehen zu lassen.

vorhandener oder sogar zur Entwicklung ganz neuer und den Forschungsgegenstand adäquater abbildender Theorien führen (Kelle/Erzberger 2004: 307, Lamnek 2005: 287).

Problematischer als der Umgang mit komplementären Ergebnissen gestaltet sich jedoch der mit divergenten (vgl. Lamnek 2005: 286-289). Da es unbefriedigend erscheint, widersprüchliche Aussagen einfach gleichwertig nebeneinander stehen zu lassen, werden verschiedene Strategien des Umgangs mit solchen Befunden eingesetzt. So wird häufig ein Ergebnis stärker gewichtet als das andere, oder es wird versucht, Ad-hoc-Erklärungen zu generieren, die eine Brücke zwischen den Befunden bilden sollen, „die aber einen methodisch und theoretisch anderen Charakter, nämlich beobachtungsfremden, haben als die empirischen Ergebnisse selbst. Ex-post-Erklärungen stellen […] implizite Theorien dar, die in ihrem Status unsicher sind“ (Lamnek 2005: 287). Möglicherweise werden (insbesondere unter dem Druck von Auftragsforschung) nicht ins Bild passende Befunde auch schlicht unterschlagen oder geben im besten Fall Anlass zu Folgestudien (ebd.: 288).

3. Formen der Triangulation nach Denzin (1970)

Ausgehend von den beschriebenen Zielsetzungen von Triangulation lassen sich verschiedene Formen unterscheiden, wobei die bekannteste Klassifikation die nach Denzin (1970: 301-310) ist:

- Daten-Triangulation/*data triangulation* (Daten, die aus verschiedenen Quellen, von verschiedenen Zeitpunkten, unterschiedlichen Orten, unterschiedlichen Personen stammen),
- Forscher-Triangulation/*investigator triangulation* (Einsatz verschiedener Datenerheber, -aufbereiter, -auswerter),
- Theorien-Triangulation/*theory triangulation* (Annäherung an den Forschungsgegenstand aus unterschiedlichen theoretischen Perspektiven),
- Methoden-Triangulation/*methodological triangulation*[2] (*within-method*, z. B. Verwendung unterschiedlicher Subskalen in einem Fragebogen; *between-/across-method*, Verwendung unterschiedlicher Methoden).

Von *Datentriangulation* kann z. B. gesprochen werden, wenn ein Beobachtungsverfahren an verschiedenen Schulen oder messwiederholend zu

2 Während Denzin selbst von *methodologischer* Triangulation spricht, was eigentlich zutreffender, da umfassender ist, hat sich im Deutschen die Übersetzung mit *Methoden*-Triangulation durchgesetzt.

mehreren, aufeinander folgenden Zeitpunkten eingesetzt wird. *Forschertriangulation* findet statt, wenn zwei oder mehrere Forscher (z. B. im Rahmen einer Forschungswerkstatt) ein Interviewtranskript diskutieren oder wenn zwei oder mehrere Forscher Lernersprache profilanalytisch analysieren und ihre Ergebnisse miteinander vergleichen. *Theorientriangulation* meint, dass zwei oder mehrere ganz unterschiedliche theoretische Zugänge zu Daten gewählt werden. Denkbar wäre im Kontext unserer Fächergruppe beispielsweise, eine aufgenommene Sprachfördersituation für das Deutsche als Zweitsprache einerseits konversationsanalytisch, andererseits linguistisch auf bildungssprachliche Merkmale und mögliche, damit verbundene Verständigungsschwierigkeiten hin zu untersuchen. Aguado (2014, vgl. auch Denzin selbst 1970: 303) zufolge kommt Theorientriangulation in der Sprachlehr- und -lernforschung jedoch quasi nicht vor, da sie mit einem hohen Aufwand für den einzelnen Forscher verbunden wäre. In interdisziplinären Forschungsverbünden wäre sie jedoch durchaus vorstellbar.

Methodentriangulation schließlich ist die wohl am häufigsten vorkommende Form der Triangulation, die bereits von Denzin (1970) zentral besprochen wird und in den Sozialwissenschaften auch unter dem Begriff der *multiplen Operationalisierung* diskutiert wird (Aguado/Riemer 2001: 247). Zu unterscheiden sind in diesem Zusammenhang Methodentriangulation innerhalb einer Methode (*within-method-triangulation*), wenn z. B. im Rahmen eines Interviews Elemente des narrativen Interviews und Elemente des Leitfadeninterviews miteinander kombiniert werden (Flick 2004: 312-313), und zwischen unterschiedlichen Methoden (*between-method-triangulation*), wenn z. B. ein reaktives Verfahren wie ein narratives Interview mit einem non-reaktiven Verfahren wie einer Dokumentenanalyse kombiniert wird (Flick 2004: 313). Letztere Form tritt dabei wiederum wesentlich häufiger als erstere auf.

Diese beiden Typen der Triangulation haben nicht nur klassifikatorische Bedeutung, sondern verweisen auch auf methodologische Differenzierungen: Die *within-method* bezieht sich im Wesentlichen auf den Kreuzvergleich der Techniken zum Zwecke der Feststellung der internen Konsistenz oder Reliabilität, während die *between-method* den Grad der externen Validität erhöhen soll (Lamnek 2005: 278).

Es lassen sich also wiederum Bezüge zwischen den weiter oben erläuterten zwei Zielsetzungen von Triangulation herstellen. Einige Beispiele sinnvoller zwischenmethodischer Triangulation für unsere Fächergruppe, die hier zur Veranschaulichung angeführt werden sollen, geben McDonough/McDonough (1997: 222):

1. Using questionnaires and selective interviewing in needs analysis.
2. Combining observation and documentary evidence in programme evaluation.
3. Using test data, experiment and self-reporting in research on reading comprehension.
4. Mixing diaries, interviews, documents, all from different sources, in case studies of individuals.
5. Student course evaluation by prompted focus group discussion at the end of the course with diaries written during the course.

Flick (2004: 315-316) führt außerdem den Begriff der *systematischen Perspektiventriangulation* als weitere Triangulationsform ein, die der Theorientriangulation ähnele, sich Flick zufolge aber auf alle vier Triangulationsformen beziehen lasse, z. B. wenn über Methodentriangulation subjektive Sichtweisen, z. B. von Beratern, mit der Beschreibung alltäglicher Routinen trianguliert würden. Dabei gehe es immer um Ergänzung, nicht um „Aufhebung von Widersprüchen", wobei Triangulation auch als Mittel zur Demokratisierung von Forschung betrachtet wird, da verschiedene Perspektiven einander gleichwertig behandelt werden sollen. Wenn beispielsweise Schüler ein innovatives Unterrichtsmaterial negativ evaluieren, dann ist das so stehenzulassen, auch wenn Eltern und Lehrer das Material hervorragend finden. Flick (2011a: 12) verwendet *Perspektiventriangulation* somit offenbar als eine Art Oberbegriff. Zu beachten ist, dass es sich bei allen beschriebenen Formen von Triangulation rein technisch betrachtet immer um die Triangulation von Daten bzw. Datensätzen handelt, da jede Triangulationsform die Existenz mindestens zweier Datensätze voraussetzt, welche zueinander in Beziehung gesetzt werden.[3] Von *Datentriangulation* wird jedoch nur gesprochen, wenn zwei Datensätze verglichen werden, die mit derselben Methode erhoben wurden (vgl. Tab. 1).

Auch die Einbettung einer Studie in ihren Forschungskontext im Rahmen von Forschungsüberblick und Diskussion könnte in einer breiteren Sicht auf erkenntnistheoretischer Ebene als eine Art von Triangulation betrachtet werden, da Bezüge zwischen bereits vorhandenen Daten und den neu erhobenen hergestellt werden. Insbesondere Replikationsstudien können in diesem Zusammenhang sehr erhellend sein.

Nicht um Triangulation handelt es sich hingegen, wenn im Rahmen derselben Studie schlicht unterschiedliche Variablen gemessen werden,

---

3 Auch im Falle von Forscher-Triangulation z. B. entstehen, wenn zwei Forscher denselben Datensatz analysieren, zwei Datensätze, nämlich die Analyse von Forscher A und die Analyse von Forscher B, welche im Anschluss miteinander verglichen respektive trianguliert werden.

wenn also z. B. mittels Fragebogen die Sprachlernmotivation von Untersuchungsteilnehmenden und mittels Test der Sprachstand derselben Personen ermittelt werden. Dann liegen zwar auch zwei Datensätze vor, die (im beschriebenen Fall z. B. in Form einer Korrelation) aufeinander bezogen werden, jedoch dienen die Messungen weder der Validierung noch der Erkenntniserweiterung, sondern operationalisieren zwei ganz unterschiedliche Forschungsgegenstände.

Tabelle 1: Triangulationsformen
(D = Datensatz, F = Forscher, M = Methode, T = Theorie)

Daten-Triangulation	**Forscher-Triangulation**	**Theorien-Triangulation**	**Methoden-Triangulation (*within-method*)**	**Methoden-Triangulation (*accross-method*)**
D1 (M1) ⇳ D2 (M1)	F1 ⬄ F2 ⬂ ⬃ D	T1 ⬄ T2 ⬂ ⬃ D	D (M1 ⬄ M2)	D1 (M1) ⇳ D2 (M2)
Perspektiven-Triangulation				

4. Triangulation innerhalb des qualitativen, innerhalb des quantitativen und zwischen dem qualitativen und quantitativen Forschungsparadigma(s)

Weiter gestaltet sich die Diskussion um Triangulation als Forschungsstrategie sehr unterschiedlich in Abhängigkeit davon, in welchem Forschungsparadigma sie angesiedelt wird. Möglich ist Triangulation innerhalb des qualitativen Forschungsparadigmas, Triangulation zwischen dem qualitativen und dem quantitativen Forschungsparadigma und Triangulation innerhalb des quantitativen Forschungsparadigmas. Letzteres findet gängig statt, wenn z. B. ein Sprachstandstest in unterschiedlichen Bundesländern normiert wird (Datentriangulation), Sprachstand mit zwei oder mehreren, einander ergänzenden Messinstrumenten erhoben wird (Methodentriangulation) usw. Diese Vorgehensweisen sind per Definition als Triangulation zu bezeichnen, werden jedoch in der Regel nicht als solche diskutiert, sondern einfach durchgeführt. Triangulation innerhalb des qualitativen Paradigmas hingegen wird in jüngster Zeit stark ins Zentrum der forschungsmethodologischen Diskussion gestellt und als Weg zu angestrebter Multiperspektivität teilweise sogar zu einer Art Gütekriterium qualitativer Forschung erhoben. Der erkenntnistheoretische Nutzen der Triangulation ist hier angesichts der Subjektivität qualitativer Forschungszugänge unmittelbar einsichtig und im Kern unstrittig.

Kontrovers hingegen wird die Verknüpfung qualitativer und quantitativer Datensätze gesehen, die häufig unter dem Schlagwort der *mixed methods*[4], manchmal auch *mixed methodologies* diskutiert wird und bereits als drittes forschungsmethodologisches Paradigma ausgerufen wurde (Tashakkorie/Teddlie 2003: ix-x, 2010: ix, Johnson/Onwuegbuzie 2004, Johnson et al. 2007, vgl. auch die umfassenden einführenden Publikationen von Plano Clark/Creswell 2008, Creswell 2009, Creswell/Plano Clark 2011). Kontrovers läuft die Debatte deshalb, weil es zwischen Vertretern des qualitativen und des quantitativen Lagers nach wie vor ungeklärt ist, in welchem erkenntnistheoretischen Verhältnis beide Paradigmen zueinander stehen. Während manche beide Herangehensweisen an empirische Forschung für gänzlich unvereinbar halten und einen „Paradigmenkrieg" führen (was mittlerweile allerdings eher eine forschungshistorische Position darstellt), propagieren andere die Eignung der jeweiligen Paradigmen in Abhängigkeit von der Forschungsfrage, wieder andere ordnen entweder die qualitative oder die quantitative Forschung der jeweils anderen Richtung über, halten den sukzessiven Einsatz qualitativer und quantitativer Forschung für zielführend oder – und das ist die genuin triangulative Position – fordern die absolut gleichberechtigte Verknüpfung beider Zugänge (vgl. Mayring 2001: 7-9, Flick 2011b: 40-42, Gläser-Zikuda et al. 2012: 8 für einen Überblick über Kombinationsmöglichkeiten qualitativer und quantitativer Ansätze auf Ebene des Forschungsdesigns).

Sehr gängig vertreten wird der Standpunkt, dass qualitative und quantitative Forschung in Abhängigkeit von der Forschungsfrage und in unterschiedlichen Phasen des Erkenntnisprozesses jeweils angemessener erscheinen. Dabei wird auf das *Phasen-* oder *Vorstudienmodell* zurückgegriffen (Barton/Lazarsfeld 1955, Mayring 2001: 8, Kelle/Erzberger 2004: 300-302, Lamnek 2005: 281-282, Gläser-Zikuda et al. 2012: 8), das postuliert, dass explorative, Hypothesen generierende Vorstudien qualitativ angelegt sein sollten, Hypothesen testende Folgestudien hingegen quantitativ. Diese Sequenzierung wird jedoch von einigen Wissenschaftlern als Abwertung qualitativer Forschung (miss-)verstanden, da qualitative Forschung der quantitativen in diesem Fall lediglich zuarbeite.[5] Im Gegenzug wird von manchen qualitativen Forschern eine

---

4 In der anglophonen Literatur hingegen wird Triangulation hingegen genau anders herum meist als eine Unterform von *mixed methods* betrachtet (vgl. die Diskussion bei Johnson et al. 2007, Bergman 2011: 272).

5 Abgesehen davon gilt die Annahme, qualitative Forschung diene der Hypothesengenerierung, quantitative Forschung der Hypothesentestung, zwar prototypisch bzw. in der großen Mehrheit der Fälle, jedoch nicht ausnahmslos. So kann eine quantitative

erkenntnistheoretische Überordnung qualitativer Forschung über die quantitative propagiert, eine Position, die „seltener, dafür aber radikaler vertreten“ (Flick 2011b: 41) wird. Argumentiert wird, dass qualitative Daten aufgrund ihrer Komplexität für sich allein genommen interpretierbar und damit fruchtbar seien, quantitative Daten hingegen häufig aus sich selbst heraus nicht erklärbar seien (vgl. Flick 2011b: 41-42). So könnte man z. B. quantitativ messen, dass eine Schulklasse höhere Leistungszuwächse zeigt als die Vergleichsklasse, wüsste aber aus der reinen Messung heraus nicht, warum dies der Fall ist. Qualitative Unterrichtsbeobachtungen, Befragungen usw. hingegen könnten hier Erklärungen liefern. Das ist sicherlich zutreffend; anders herum kann man aber natürlich auch argumentieren, dass die Leistungsunterschiede, die interpretiert werden sollen, in der rein qualitativen Betrachtung beider Schulklassen gar nicht sichtbar geworden wären. Beide Zugänge tragen somit nur gemeinsam zu einem vollständigeren Bild des Untersuchungsgegenstandes bei (vgl. Settinieri 2012: 250-253).

Dies führt Flick dazu, im Rahmen triangulativer Ansätze eine echte Gleichberechtigung beider Paradigmen zu fordern und den Begriff der Triangulation auf Studien zu beschränken, die diese gewährleisten:

> Es gibt nur wenige Beispiele, in denen Methoden entwickelt werden, die tatsächlich qualitative und quantitative Strategien in einer Methode integrieren. Viele Fragebogen enthalten offene oder Freitext-Fragen, was in manchen Kontexten schon als qualitative Forschung verstanden wird, obwohl kaum ein methodisches Prinzip qualitativer Forschung mit diesen Fragen aufgegriffen wird. Dies ist weniger eine explizite Kombination beider Formen der Forschung als der Versuch, einen Trend aufzugreifen. (Flick 2011b: 48)

Tashakkori/Teddlie (2003: xi) gehen sogar so weit zu fordern, verschiedene Ansätze in allen Untersuchungsphasen zu berücksichtigen und eine Transformation der Daten zwischen den Ansätzen vorzunehmen. Dieser Maximalanforderung an Triangulation steht die Minimalforderung, Triangulation setze voraus, dass sich unterschiedliche Datensätze auf einen gemeinsamen Untersuchungsgegenstand beziehen, gegenüber. Eine vermittelnde Position fordert die Ausschöpfung beider Zugänge in ihrer Eigengesetzlichkeit und eine Gleichgewichtung insbesondere bei der Dateninterpretation. Sie integriert damit auch die Forderung nach der Gegenstandsangemessenheit empirischer Vorgehensweisen

explorative Faktorenanalyse beispielsweise durchaus zur Hypothesengenerierung eingesetzt werden.

und zielt darauf, die Stärken beider Paradigmen zu nutzen. Eine Transformation zwischen beiden Paradigmen erscheint hingegen problematisch. So sollten beispielsweise Daten aus narrativen Interviews zusätzlich zu ihrer qualitativen Analyse nicht ohne Weiteres auch quantitativ ausgewertet werden, da die Daten nicht nach Gütestandards quantitativer Forschung erhoben wurden. Aussagen wie *Sieben von zehn Interviewten sind der Ansicht, dass…* oder gar *70% der Interviewten…* sind daher irreführend. Und genuin quantitative Daten, wie beispielsweise Angaben zum Umfang erhaltenen Sprachunterrichts, qualitativ zu analysieren, ist häufig gar nicht möglich, da den Daten dazu in der Regel die Tiefe fehlt.

Anstelle dogmatischer Forderungen erscheint zusammenfassend eine stärkere Berücksichtigung forschungspragmatischer Aspekte zielführender zu sein. Die Über- bzw. Unterordnung des einen oder des anderen Zugangs kann dabei manchmal durchaus gegenstandsangemessen und der Beantwortung der Forschungsfrage zuträglich sein (vgl. auch die Beispiele bei McDonough/McDonough 1997: 223-224). Wichtig ist vielmehr, dass grundsätzlich das Potenzial beider Zugänge mit Blick auf den konkreten Untersuchungsgegenstand erkannt und genutzt wird.

5. Typologie des Triangulationsbegriffs

Neben den bereits in Abschnitt 2. bis 4. erläuterten Kriterien können Triangulationsformen zusätzlich nach weiteren Aspekten klassifiziert werden. Ebenfalls sehr gängig ist die Differenzierung in Vorstudien-, Verallgemeinerungs-, Vertiefungs- und Triangulationsmodell (Mayring 2001: 7-9). Während im Rahmen des Vorstudienmodells eine qualitative Studie lediglich zur Hypothesengenerierung im Vorfeld einer quantitativen Studie dient, beschreibt das Verallgemeinerungsmodell den Fall einer für sich stehenden qualitativen Studie, deren Ergebnisse durch eine quantitative Folgestudie zusätzlich auf eine Population übertragen werden sollen. Im Vertiefungsmodell werden Ergebnisse einer vorangehenden quantitativen Studie mittels qualitativer Folgestudie interpretativ verdichtet, und im Triangulationsmodell schließlich werden qualitative und quantitative Methoden sowohl gleichzeitig als auch gleichwertig eingesetzt, um zu einem Gesamtergebnis zusammengeführt zu werden. „Im ersten Fall ist der Stellenwert bzw. die Funktion qualitativer Methoden überwiegend vorläufig-explorativer, im zweiten eigenständig deskriptiver, im dritten nachträglich-profundierender und im vierten gleichzeitig-synergetischer Art […].“ (Lamnek 2005: 281)

Weiter kann die Verknüpfung von Daten grundsätzlich über Einzelfälle oder über Gesamtdatensätze erfolgen, wobei Flick (2004: 316) die Triangulation am

Einzelfall als die „konsequenteste Variante“ bezeichnet. Sie liegt beispielsweise dann vor, wenn aus 200 Probanden einer Fragebogenstudie 10 für zusätzliche vertiefende Interviews ausgewählt werden (vgl. Flick 2011b: 44). Ein klarer Vorteil dieser Vorgehensweise liegt darin, dass sie „Vergleiche und Verbindungen auch auf höherer Ebene“ ermöglicht; nachteilig hingegen erscheint die vergleichsweise hohe Belastung einzelner Teilnehmer, die die Gefahr von Ausfällen deutlich erhöhen kann (Flick 2004: 316-317).

Von einer Verknüpfung über Datensätze hingegen spricht man, wenn beispielsweise aus Interviewdaten eine Typologie abgeleitet wird, die mit der Verteilung korrespondierender Fragebogenantworten in Zusammenhang gebracht wird (Flick 2011b: 44-45). Diese Vorgehensweise ist in der Regel weniger aufwändig als die Triangulation am Einzelfall, allerdings insofern problematisch, als dass die Samples nicht uneingeschränkt miteinander vergleichbar sind. Ebenfalls problematisch für die Vergleichbarkeit von Datensätzen kann in diesem Zusammenhang auch eine Zeitversetzung zwischen ihnen sein (Flick 2004: 317).

Schließlich kann man, in Abhängigkeit von der theoretischen Fundierung und dem Reflexionsniveau triangulativer Vorgehensweisen, explizite und implizite Triangulation einander gegenüberstellen. Wenn beispielsweise verschiedene Methoden, wie insbesondere in ethnografischer Forschung gängig, eher spontan in der Erhebungssituation (z. B. im Rahmen einer längeren teilnehmenden Beobachtung) eingesetzt werden, spricht Flick (2004: 314) von *impliziter* Triangulation. Diese grenzt sich von im Vorfeld einer Erhebung konkret geplanter expliziter Triangulation ab. Einen zusammenfassenden Überblick über Triangulationsformen bietet Tab. 2 (vgl. auch Mayring 2001, Creswell et al. 2003, Johnson/Onwuegbuzie 2004: 22, Dörnyei 2007: 169-173).

6. Praxisprobleme

In der Forschungspraxis ergeben sich einige Schwierigkeiten beim Einsatz triangulativer Strategien, insbesondere im Falle der Verknüpfung quantitativer und qualitativer Daten. Bei einer Triangulation über den Einzelfall muss hier bedacht werden, dass sich Sampling-Entscheidungen sowohl an quantitativen als auch an qualitativen Standards orientieren müssen. Während für die quantitative Teilstudie geklärt werden muss, wie viele Untersuchungsteilnehmer ausgewählt werden müssen (z. B. mittels statistischer Power-Analyse) und wie diese so ausgewählt werden können, dass die Stichprobe als repräsentativ für die untersuchte Population gelten kann, ist für die qualitative

Teilstudie (häufig auch sukzessive) zu entscheiden, welche Untersuchungsteilnehmer für die jeweilige Fragestellung interessant sein könnten, weil sie z. B. als typisch oder als kontrastierend gelten können (vgl. zu Samplingstrategien genauer Morse/Niehaus 2009: 63-76, Teddlie/Tashakkori 2009: 168-196, Settinieri 2014). Die zentrale Anforderung besteht hier also wiederum darin, beide Paradigmen in ihrer Eigengesetzlichkeit zu achten (vgl. z. B. Steinke 2004 zu Gütekriterien qualitativer Forschung, Bortz/Döring 2006: 194-206 zu Gütekriterien quantitativer Forschung, Riemer/Settinieri 2010: 765-771 zu einer Gegenüberstellung des quantitativen und qualitativen Forschungsparadigmas, Mayring 2012 zu Forschungsstandards für *mixed methods designs*). Sollen beispielsweise im Rahmen eines Vertiefungsmodells nach einer quantitativen Fragebogenstudie mit einer auf Grundlage der Fragebogendaten ausgewählten kleineren Gruppe von Befragten zusätzliche Interviews durchgeführt werden, um vorgefundene quantitative Befunde besser verstehen zu können, so muss die Auswahl dieser Personen entlang der Forschungsfrage nach qualitativen Kriterien erfolgen.

Das wohl größte Problem triangulativer Studien besteht jedoch darin, angesichts in kürzester Zeit entstehender „Datenberge" Strategien der Reduktion und der strukturierten Ergebnisdarstellung zu entwickeln. Grundsätzlich ist dabei hilfreich, sich immer wieder auf seine genaue Forschungsfrage zurückzubesinnen, um anhand derer fokussieren und selektieren zu können. In der Regel werden die einzelnen Datensätze dann in der Publikation zunächst unabhängig voneinander ausgewertet und interpretiert, um sie in ihrem Erhebungskontext und in ihrer Gänze zu erfassen, und erst in einem zweiten Schritt zueinander in Beziehung gesetzt (vgl. Aguado/Riemer 2001: 253). Bei sequenziellen Designs versteht sich die getrennte Darstellung ohnehin von selbst. In einigen Fällen kann aber auch eine unmittelbar integrierte Auswertung sinnvoll sein, insbesondere bei der Triangulation über Einzelfälle (vgl. genauer Teddlie/Tashakkori 2009: 249-314 zu Verfahren der Datenanalyse und -interpretation im Rahmen von *mixed methods designs*). Spätestens bei der Gesamtinterpretation der Ergebnisse einer Studie ist die integrierte Darstellung triangulativer Elemente allerdings unabdingbar, um überhaupt von Triangulation sprechen zu können.

Tabelle 2: Formen von Triangulation

Zielsetzung	Ebene	Forschungs-phase	Para-digmen	Abfolge	Gewichtung	Modell	Verknüp-fungse bene	Theoretische Fundierung
Validierung	Daten	Datenerhebung	QN – QN	QL vor QN	höhere Relevanz der QN	Vorstu-dienmo-dell	Datensatz	implizit
Erkennt-niserweite-rung	Forscher	Datenaufbereitung	QL – QL	QN vor QL	höhere Relevanz der QL	Verallge-meine-rungs-modell	Einzelfall	explizit
	Theorien	Datenanalyse	QN – QL	Gleichzei-tigkeit	Gleich berechtigung	Vertie-fungs-modell		
	Methoden	Dateninterpretation				Triangu-lations-modell		
	Perspektiven							

7. Fazit

Zusammenfassend ist Triangulation die Bezugnahme unterschiedlicher Datensätze auf denselben Forschungsgegenstand, wobei qualitative und quantitative Ansätze einander ergänzende, gleichwertige Potenziale bieten und in ihrer Eigengesetzlichkeit geachtet werden sollten. In rein qualitativer Forschung ist Triangulation in den letzten Jahren ein Güteindikator geworden und auf dem Wege zur *conditio sine qua non*. Die Verknüpfung qualitativer und quantitativer Ansätze in einer Forschungsarbeit bedarf hingegen stets einer expliziten theoretischen Begründung und erscheint lediglich auf Ebene der Forscher-Triangulation in jedem Fall gewinnbringend. Denn unabhängig vom Forschungsparadigma stellt es immer ein Risiko dar, eine Studie zu veröffentlichen, ohne dass mehr als eine Person die Daten (zumindest in Ausschnitten) analysiert und interpretiert hätten.

Dreh- und Angelpunkt der Diskussion um Sinn und Unsinn von Triangulation sind die zwei bereits ausführlicher angesprochenen Aspekte der echten Integration der unterschiedlichen Perspektiven sowie die Forderung nach einer zumindest potenziell gleichberechtigten Rolle quantitativer und qualitativer Forschungszugänge. Dieses Prinzip der methodologischen Offenheit (Lamnek 2005: 278-279) setzt auf Seiten der Forschenden jedoch forschungsmethodologische Kompetenzen im Bereich beider Paradigmen voraus, die nicht selbstverständlich als gegeben gelten können. Forschungsverbünde scheinen hier praktikabler als Einzelstudien zu sein (z.B. Johnson/Onwuegbuzie 2004: 21, Dörnyei 2007: 46-47, 174-175, Bergman 2011: 274). Wichtig ist in diesem Zusammenhang auch, die Reflexion über das Zusammenspiel triangulativer Elemente in einer Studie transparent und für andere Forscher nachvollziehbar zu gestalten, damit diese sich selbst ein Bild von der Aussagekraft der Forschungsergebnisse machen können (vgl. Aguado 2000, Bärenfänger/Stevener 2001).

Eine solche Triangulation kann dann entweder der Validierung oder der Erkenntniserweiterung dienen, wobei insbesondere die zweite Zielsetzung für unsere Fächergruppe relevant ist. Mit der Multiplikation von Perspektiven und somit von Datensätzen geht jedoch stets auch ein Ansteigen des Zeit- und Kostenbedarfs einher (vgl. Lamnek 2005: 291, Bergman 2011: 274), der nicht zu unterschätzen und in die Planung einer triangulativen Studie miteinzubeziehen ist. Gerade im Falle von Qualifikationsarbeiten, die von einer Person allein durchgeführt werden sollen, entstehen sonst leicht „Datenfriedhöfe" (Riemer 2010: 430).

Und schließlich kann eine „[m]ultimethodische Vorgehensweise […] ebenso wie unimethodisches Vorgehen dann nutzlos werden, wenn mit ihrer Hilfe

die falsche Fragestellung analysiert wird“ (Lamnek 2005: 290) oder auch bereits wenn nicht jede einzelne Methode für sich genommen angemessen gewählt wird (vgl. Flick 2011b: 54). Lamnek (2005: 290-291) spricht in diesem Zusammenhang von Triangulation als einer „forschungs- und finanzierungstaktischen Modeerscheinung ohne ernsthafte wissenschaftliche Überzeugung“ und als „normative[r] Leerformel“ (Lamnek 2005: 279). In diesem Sinne sollte der Titel des vorliegenden Beitrags vermutlich umformuliert werden in die Frage: *Triangulierst Du noch, oder forschst Du schon?*[6]

Literatur

Aguado, Karin (2000). Empirische Fremdsprachenerwerbsforschung. Ein Plädoyer für mehr Transparenz. In: Aguado, Karin (Hrsg.). *Zur Methodologie in der empirischen Fremdsprachenerwerbsforschung.* Baltmannsweiler: Schneider Verlag Hohengehren, 119-132.

Aguado, Karin (2014). Triangulation. In: Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hrsg.). *Einführung in empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache.* Tübingen: UTB, 47-56.

Aguado, Karin / Riemer, Claudia (2001). Triangulation: Chancen und Grenzen mehrmethodischer empirischer Forschung. In: Aguado, Karin / Riemer, Claudia (Hrsg.). *Wege und Ziele. Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen).* Baltmannsweiler: Schneider Verlag Hohengehren, 245-257.

Bärenfänger, Olaf / Stevener, Jan (2001). Datenerhebungsverfahren und ihre Evaluation. Ein Kriterienkatalog. *Fremdsprachen Lehren und Lehren* 30, 13-27.

Barton, Allen H. / Lazarsfeld, Paul F. ([1955] 1979). Einige Funktionen von qualitativer Analyse in der Sozialforschung. In: Hopf, Christel / Weingarten, Elmar (Hrsg.). *Qualitative Sozialforschung.* Stuttgart: Klett-Cotta, 41-89.

Bergman, Manfred Max (2011). The Good, the Bad, and the Ugly in Mixed Methods Research and Design. *Journal of Mixed Methods Research* 5, 271-275.

Bortz, Jürgen / Döring, Nicola (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* 4., Aufl. Heidelberg: Springer.

Brown, James Dean / Rodgers, Theodore S. (2002). *Doing Second Language Research.* Oxford: Oxford University Press.

6 Ich danke Sevilen Demirkaya und Nazan Gültekin-Karakoç herzlich für zahlreiche Diskussionen und hilfreiche Hinweise zu früheren Versionen dieses Aufsatzes.

Campbell, Donald T. / Fiske, Donald W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin* 56, 81-105.

Creswell, John W. (2009). *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches.* 3. Aufl. Thousand Oaks u.a.: Sage.

Creswell, John W. / Plano Clark, Vicki L. (2011). *Designing and Conducting Mixed Methods Research.* 2. Aufl. Thousand Oaks u.a.: Sage.

Creswell, John W. / Plano Clark, Vicki L. / Gutmann, Michelle L. / Hanson, William E. (2003). Advanced Mixed Methods Research Design. In: Tashakkori, Abbas / Teddlie, Charles (Hrsg.). *Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks u.a.: Sage, 209-240.

Denzin, Norman K. (1970). *The Research Act. A Theoretical Introduction to Sociological Methods.* Chicago: Aldine.

Denzin, Norman K. / Lincoln, Yvonna S. (1994). Introduction: Entering the Field of Qualitative Research. In: Denzin, Norman K. / Lincoln, Yvonna S. (Hrsg.). *Handbook of Qualitative Research.* Thousand Oaks u.a.: Sage, 1-17.

Dörnyei, Zoltán (2007). *Research Methods in Applied Linguistics.* Oxford: Oxford University Press.

Fielding, Nigel G. / Fielding, Jane L. (1986). *Linking Data.* Beverly Hills: Sage.

Flick, Uwe (2004). Triangulation in der qualitativen Forschung. In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch.* 3. Aufl. Reinbek bei Hamburg: Rowohlt, 309-318.

Flick, Uwe (2011a). *Triangulation. Eine Einführung.* 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick, Uwe (2011b). *Qualitative Sozialforschung. Eine Einführung.* 4., vollst. überarb. und erw. Neuausgabe. Reinbek bei Hamburg: Rowohlt.

Gläser-Zikuda, Michaela / Seidel, Tina/Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha (2012). Mixed Methods in der empirischen Bildungsforschung – eine Einführung in die Thematik. In: Gläser-Zikuda, Michaela / Seidel, Tina / Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha (Hrsg.). *Mixed Methods in der empirischen Bildungsforschung.* Münster u. a.: Waxmann, 7-13.

ILMES – *Internet-Lexikon der Methoden der empirischen Sozialforschung* (o. J.). Online: http://psydok.sulb.uni-saarland.de/volltexte/2004/260/html/ilmes.htm [4.7.2014].

Johnson, Burke R. / Onwuegbuzie, Antony J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher* 33, 14-26.

Johnson, Burke R. / Onwuegbuzie, Antony J. / Turner, Lisa A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research* 1, 112-133.

Kelle, Udo / Erzberger, Christian (2004). Qualitative und quantitative Methoden: kein Gegensatz. In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch.* 3. Aufl. Reinbek bei Hamburg: Rowohlt, 299-309.

Lamnek, Siegfried (2005). *Qualitative Sozialforschung. Lehrbuch.* 4., Aufl. Weinheim/Basel: Beltz.

Mayring, Philipp (2001). Kombination und Integration qualitativer und quantitativer Analyse. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* 2, Art. 6. Online: http://nbn-resolving.de/urn:nbn:de:0114-fqs010162 [10.3.2014].

Mayring, Phillip (2012). Mixed Methods – ein Plädoyer für gemeinsame Forschungsstandards qualitativer und quantitativer Methoden. In: Gläser-Zikuda, Michaela / Seidel, Tina / Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha (Hrsg.). *Mixed Methods in der empirischen Bildungsforschung.* Münster u. a.: Waxmann, 287-300.

McDonough, Jo / McDonough, Steven (1997). *Research Methods for English Language Teachers.* London: Arnold.

Morse, Janice M. / Niehaus, Linda (2009). *Mixed Method Design. Principles and Procedures.* Walnut Creek: Left Coast Press.

Plano Clark, Vicki L. / Creswell, John W. (Hrsg.) (2008). *The Mixed Methods Reader.* Los Angeles u.a.: Sage.

Riemer, Claudia (2010). Erste Schritte empirischer Forschung: Themenfindung, Forschungsplanung, forschungsmethodologische Entscheidungen. In: Chlosta, Christoph / Jung, Matthias (Hrsg.). *DaF integriert: Literatur – Medien – Ausbildung. Tagungsband der 36. Jahrestagung des Fachverbandes Deutsch als Fremdsprache 2008.* Göttingen: Universitätsverlag, 423-434.

Riemer, Claudia (2011). Empirie und Fremdsprachenforschung: Herausforderung Forschungsmethodik. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.). *Fremdsprachen lehren und lernen: Rück- und Ausblick. Arbeitspapiere der 30. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts.* Tübingen: Narr, 194-203.

Riemer, Claudia / Settinieri, Julia (2010). Empirische Forschungsmethoden in der Zweit- und Fremdsprachenerwerbsforschung. In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia (Hrsg.). *Deutsch als Fremd- und Zweitsprache.* Berlin: de Gruyter, 764-781.

Settinieri, Julia (2012). Statistische Verfahren: Grundlagenbeitrag. In: Doff, Sabine (Hrsg.). *Fremdsprachenunterricht empirisch erforschen. Grundlagen – Methoden – Anwendung.* Tübingen: Narr, 249-270.

Settinieri, Julia (2014). Planung einer empirischen Studie. In: Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hrsg.). *Einführung in empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache.* UTB, 57- 71.

Steinke, Ines (2004). Gütekriterien qualitativer Forschung: In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch.* 3. Aufl. Reinbek bei Hamburg: Rowohlt, 319-331.

Tashakkori, Abbas / Teddlie, Charles (2003). Preface. In: Tashakkori, Abbas / Teddlie, Charles (Hrsg.): *SAGE Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks u.a.: Sage, ix-xv.

Tashakkori, Abbas / Teddlie, Charles (2010). Preface. In: Tashakkori, Abbas / Teddlie, Charles (Hrsg.): *SAGE Handbook of Mixed Methods in Social and Behavioral Research.* 2. Aufl. Thousand Oaks u.a.: Sage, ix-xv.

Teddlie, Charles / Tashakkori, Abbas (2009). *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences.* Thousand Oaks u.a.: Sage.

Webb, Eugene J. / Campbell, Donald T. / Schwartz, Richard D. / Sechrest, Lee (1966). *Unobtrusive Measures. Nonreactive Research in the Social Sciences.* Chicago: Rand McNally.

Wie kann Interkulturelle Kommunikative Kompetenz im Englischunterricht der Grundschule untersucht werden? – Begründung eines Forschungsdesigns

Sonja Brunsmeier

Kinder wachsen in einer multikulturellen und multilingualen Welt auf. Sowohl in ihrem Alltag als auch im Klassenzimmer treffen Grundschulkinder auf verschiedene Kulturen und Sprachen. Die Beschäftigung mit fremden Sprachen und Kulturen ist somit Teil der kindlichen Lebenserfahrung. Für einen kompetenten Umgang mit dieser Vielfalt ist Interkulturelle Kommunikative Kompetenz (IKK) unabdingbar. Aktuell liegen für diese spezifische Gruppe jedoch weder ein überzeugendes theoretisches Konzept noch konkrete Umsetzungsideen vor. Das Ziel dieser Studie ist es, das Konstrukt IKK für den Englischunterricht in der Grundschule zu erfassen und über die Entwicklung und Erprobung geeigneter Aufgaben zu konkretisieren und empirisch zu überprüfen. Der vorliegende Beitrag begründet das methodische Vorgehen der Studie.

1. Die Relevanz der Interkulturellen Kommunikativen Kompetenz im Englischunterricht der Grundschule

Kinder wachsen in einer multikulturellen und multilingualen Welt auf, in der die englische Sprache durch ihre globale Verbreitung eine besondere Rolle spielt. Sowohl in ihrem Alltag als auch im Klassenzimmer treffen Grundschulkinder auf verschiedene Kulturen und Sprachen. Die Beschäftigung mit fremden Sprachen und Kulturen ist somit Teil der kindlichen Lebenserfahrung. Durch Medien und in der Alltagssprache treffen Kinder verstärkt auf Anglizismen, welche die Rolle der englischen Sprache als *Lingua Franca* unterstreichen. Vor diesem Hintergrund kommt dem Englischunterricht in der Primarstufe bei der Vermittlung von Interkultureller Kommunikativer Kompetenz (IKK) eine besondere Bedeutung zu.

> Eine frühzeitige Entwicklung der fremdsprachlichen Kompetenzen wird als Vorbereitung auf das Leben in einer durch sprachliche und kulturelle Vielfalt geprägten Umwelt als wichtig erachtet. Aus diesem Grund sollen alle Kinder Fremdsprachen über einen längeren Zeitraum hinweg und in größerer Vielfalt lernen. Zugleich sollen sie durch die Erfahrung und das Verstehen von Fremdem eine erweiterte Perspektive auf ihre Umwelt und die eigene Identität gewinnen. In dem kontinuierlichen Ausbau und Zusammenwirken dieser beiden Dimensionen wird insbesondere eine Voraussetzung für die erfolgreiche

> Kommunikation und Vermittlung in interkulturellen Kontexten gesehen. Dem fremdsprachlichen Lernen in der Grundschule kommt damit die Aufgabe zu, den Grundstein für diese interkulturelle kommunikative Kompetenz zu legen (Marschollek 2011: 455).

Die Beherrschung der englischen Sprache wird auch als vierte Kulturtechnik (vgl. Brusch 2000: 34, Pfitzer 2006: 33) bezeichnet, da sie Kinder zu „[k]ulturelle[r] Teilhabe“ (Pfitzer 2006: 27) befähigt. Durch das Erlernen der *Lingua Franca* werden Kinder bereits im Grundschulalter auf gesellschaftliche Partizipation und den Dialog zwischen verschiedenen Kulturen vorbereitet. Durch die Auseinandersetzung mit der englischen Sprache erfahren die Grundschüler/innen gleichzeitig auch, dass Sprache „Träger kultureller Inhalte“ (Schocker-von Ditfurth 2004: 217) ist und es ihnen ermöglicht, mit anderen Kulturen in Kontakt zu treten und sich diese zu erschließen bzw. Gemeinsamkeiten und Unterschiede zwischen der eigenen und den anderen Kulturen zu entdecken (vgl. Legutke et al. 2009: 15). „[D]as Lernen von Sprachen [ist] ein durch nichts anderes zu ersetzender Anlass zum interkulturellen Lernen und zu einer seiner Grundlagen, dem Fremdverstehen“ (Christ 2007: 52). Der Englischunterricht möchte die kindliche Neugierde für das Fremde wecken. Im „geschützten Raum“ (ebd.: 52) des Klassenzimmers können andere Kulturen entdeckt und erkundet werden. Gleichzeitig wird so die Basis für eine grundlegende Offenheit gegenüber Sprachen und Kulturen geschaffen (vgl. Marschollek 2002: 119ff.). Dabei versteht sich das heutige Klassenzimmer als ein offener und gelegentlich virtueller Raum, in den viel ‚Welt‘ und ‚Erfahrung‘ eindringen (vgl. Christ 2007: 55).

Auch bildungspolitische Papiere betonen die Notwendigkeit, Lerner auf unsere multikulturelle und mehrsprachige Welt vorzubereiten. Der Gemeinsame Europäische Referenzrahmen nennt Interkulturelle Kommunikative Kompetenz (IKK) als die zentrale Kompetenz des Fremdsprachenunterrichts (vgl. Council of Europe 2001: 103ff.). Die bundeslandspezifischen Bildungspläne äußern sich entsprechend. Beispielsweise nennt der baden-württembergische Bildungsplan (vgl. Ministerium für Kultus, Jugend und Sport Baden-Württemberg 2004: 74ff.) IKK explizit als Zielvorgabe für den Englischunterricht. Gleiches gilt auch für die hessischen Bildungsstandards und Inhaltsfelder für Moderne Fremdsprachen für die Grundschule (vgl. Hessisches Kulturministerium 2010) und den hessischen Leitfaden für Moderne Fremdsprachen in der Grundschule (vgl. Institut für Qualitätsentwicklung 2011). Daraus folgt, dass die Anbahnung von IKK im Englischunterricht als gleichwertiges Ziel neben die sprachlichen Kompetenzen tritt. Die Zielformulierungen der Lehrpläne schließen den Primarbereich mit ein. Mit Beginn des Schuleintritts muss die Vermittlung von IKK fester Bestandteil der Unterrichtskultur sein. „If it is the purpose of all education to help the individual to function well in his or her society and if this society is multicultural already, then such help cannot be postponed to

the secondary level of education but must be given from the earliest stage of formal education, i.e. from the primary school onwards“ (Doyé 1999: 25).

Wenngleich Einigkeit hinsichtlich der grundlegenden Zielvorstellungen des Englischunterrichts in der Primarstufe besteht, und IKK als didaktischer Schlüsselbegriff gelten kann, findet sich keine allgemein geteilte Definition des Begriffs: „the concept remains as fuzzy as ever“ (Hu/Byram 2009: XI) und „the term has [not] been stable and each of the two terms in the phrase ‚intercultural competence‘ is likewise much debated“ (Guilherme 2013: 346). So gibt es bis heute verschiedene Vorstellungen, welche Dimensionen IKK beinhaltet. Beispielhaft seien hier Byrams Kompetenzmodell (Byram 1997) und Bennetts Stufenmodell (Bennett 1986) genannt. Beide Konzeptionen haben die fremdsprachendidaktische Diskussion der letzten Jahre maßgeblich geprägt, zeigen exemplarisch aber eben auch auf, wie unterschiedlich die Vorstellungen in Bezug auf IKK sind. Byrams (1997: 31) *model for intercultural communicative competence* bezieht sich explizit auf Bildungskontexte und Bildungsziele (vgl. ebd.: 70). Er benennt fünf gleichwertige Dimensionen (*attitudes, knowledge, skills of interpreting and relating, skills of discovery and interaction* und *critical cultural awareness*), die zentral für eine erfolgreiche interkulturelle Kommunikation sind und sich wechselseitig beeinflussen. Im Verlauf des Lernprozesses sollen die Lerner zu *intercultural speakers* (ebd.: 71) ausgebildet werden, die in der Lage sind

> to interact with people from another country and culture in a foreign language. They are able to negotiate a mode of communication and interaction which is satisfactory to themselves and the other and they are able to act as a mediator between people of different cultural origins. Their knowledge of another culture is linked to their language competence through their ability to use language appropriately – sociolinguistic and discourse competence – and their awareness of the specific meanings, values and connotations of the language. They also have a basis for acquiring new languages and cultural understandings as a consequence of the skills they have acquired in the first (ebd.).

Bennetts *Developmental Model of Intercultural Sensitivity* (1986) geht davon aus, „that as one's experience of cultural difference becomes more complex und sophisticated, one's competence in intercultural relations increases“ (Hammer/Bennett 2001: 12). In seinem sechsstufigen Modell spiegelt jede Stufe ein bestimmtes kognitives Niveau wider, welches sich in der Entwicklung bestimmter Haltungen, Einstellungen und Verhaltensweisen bezogen auf kulturelle Aspekte zeigt.

Beide Modellierungen fokussieren den Sekundarbereich und die Erwachsenenbildung. Da sich Kinder im Grundschulalter „hinsichtlich ihrer sprachlichen, intellektuellen, sozialen und psychischen Entwicklung sehr“ (Christ 2003: 449) von älteren Lernern unterscheiden, muss dem Kontext eine besondere Beachtung geschenkt werden und bestehende Modelle auf deren Übertragbarkeit überprüft

werden. Die Forschung bezieht sich „aktuell konsequenterweise vor allem auf die Frage, wie interkulturelle Lernprozesse im Kontext von Fremdsprachenunterricht empirisch rekonstruierbar, wie Entwicklungsverläufe in diesem Kontext abbildbar und mit welchen Instrumenten interkulturelle Kompetenzen messbar bzw. evaluierbar sind. […] Gleichzeitig ist es wichtig, für die Unterrichtspraxis Aufgaben zu entwickeln, die Interkulturelle Kommunikative Kompetenzen systematisch entwickeln und beschreibbar machen“ (Hu 2010: 77f.). Lehrkräfte fragen sich: „How do I teach the intercultural dimension […]?“ (Byram et al. 2002: 10). Da Aufgaben als „'environment' or 'ecosystem'“ (Cameron 2001: 21) gesehen werden, „in which the growth of skills in the foreign language takes place“ (ebd.), ist es von zentraler Wichtigkeit, Aufgabenformate für die Anbahnung von IKK in der Grundschule zu entwickeln, die es Lehrkräften ermöglichen, die ermittelten Dimensionen von IKK auf Klassenzimmerebene zu vermitteln. Freitag-Hild (2010) legt eine Aufgabentypologie für den Sekundarbereich vor und Müller-Hartmann et al. (2013: 115ff.) präsentieren Aufgabenkriterien, die geeignet sind, IKK ebenfalls für den Sekundarbereich zu vermitteln. Für den Primarbereich stehen eine Entwicklung von Aufgabenformaten und deren empirische Überprüfung noch aus.

Aus den bisherigen Überlegungen ergibt sich folgende Problemstellung, welche die Grundlage für die Studie bildet: IKK ist als ein Ziel im Fremdsprachenfrühbeginn ausgeschrieben. Durch die enge Verbindung von Sprache und Kultur wird der Englischunterricht als geeigneter Ort gesehen, um den Grundstein für diese Kompetenz bei den jungen Lernern zu legen. Jedoch ist weder die Terminologie IKK für diesen Kontext eindeutig definiert, noch stehen umfassende konkrete Umsetzungsideen (Aufgaben) für die Unterrichtspraxis im Englischunterricht der Grundschule bereit.

2. Ziel der Arbeit

Zunächst steht die Klärung der Terminologie von IKK für den Englischunterricht der Grundschule im Vordergrund. Um die Terminologie für den spezifischen Kontext Grundschule klären zu können, werden bestehende Definitionen und Modelle auch auf die Rahmenbedingungen, unter denen frühes Fremdsprachenlernen stattfindet, bezogen. Weiterhin werden vorhandene Erkenntnisse der Lern- und Entwicklungspsychologie im Kindesalter beachtet, da diese bei der Anbahnung von IKK eine wichtige Rolle spielen. So kann ermittelt werden, welche Zielvorstellungen angemessen für das Grundschulalter sind.

Basierend auf den gewonnenen Erkenntnissen soll in einem nächsten Schritt geklärt werden, wie Lernaufgaben beschaffen sein müssen, damit sie IKK bei jungen Lernern anbahnen. Dabei müssen Sprachstand, Sprachentwicklung und

Motivationslage der Kinder (vgl. Legutke et al. 2010: 292) berücksichtigt werden. Theoriegeleitet werden Aufgabendeskriptoren für diese spezifische Altersgruppe formuliert. Diese sind die Grundlage für die modellhafte Entwicklung von Aufgaben und Materialien. Ziel ist es also, altersgerechte Aufgaben und Materialien zu entwickeln, die die Anbahnung von IKK im Primarbereich ermöglichen. Abschließend werden die Aufgaben in vier Grundschulklassen empirisch erprobt und ihr Potenzial anhand der Aufgabendeskriptoren bewertet.

Aus den empirisch erprobten Aufgabenformaten können dann Rückschlüsse auf die Dimensionen, die IKK anbahnen, gezogen werden. Möglichkeiten und Grenzen der Entwicklung von IKK im Englischunterricht der Primarstufe können so dargestellt und diskutiert werden.

Abb. 1: Überblick über den Ablauf der Untersuchung

Vorstudie Rahmenbedingungen und Wissensbestände			
Literatur-recherche	Dokumenten-analyse (Bildungspläne)	Lehrwerks-analyse	Problemzentrierte Leitfadeninterviews: Grundschullehrkräfte

⇩

Modellierung von Byrams 5 Dimensionen für den Englischunterricht der Primarstufe
(Can-Do Statements)
Daraus: Generierung der Aufgabenkriterien



Hauptstudie
Empirische Überprüfung des Aufgabenpotenzials

Fallstudien:
Empirische Überprüfung der Modellaufgaben durch *Participatory Action Research*

Datensätze:
Videodaten / Lehrerbezogene Datensätze (Gespräche, Beobachtungsbögen) / Schülerbezogene Datensätze (Unterrichtserzeugnisse, Fragebögen, Interviews)

⇩

Ermittlung des Aufgabenpotenzials



Ergebnis
Empirisch überprüfte Modellaufgaben zur Entwicklung von IKK im Englischunterricht der Primarstufe

3. Methodisches Vorgehen

Die einführenden Absätze haben deutlich gemacht, dass es zu IKK kaum praktische Umsetzungen (z.B. Aufgaben) im Englischunterricht der Grundschule gibt. Es liegen keine empirisch überprüften Erkenntnisse über Möglichkeiten und Grenzen der Implementation vor. Ein quantitatives Vorgehen, welches hypothesenüberprüfend und erklärend vorgeht (vgl. Lamnek 2005: 243), ist deshalb als forschungsmethodischer Ansatz ungeeignet. Nur ein qualitatives Design ermöglicht es, sich der Thematik induktiv-verstehend zu nähern, das Feld explorativ-interpretativ zu erforschen (vgl. Caspari et al. 2003: 499ff.) und neue Theorien zu generieren (vgl. Croker 2009: 9). Abbildung 1 visualisiert das methodische Vorgehen der Forschungsarbeit.

3.1 Die Vorstudie

Literaturrecherche und Dokumentenanalyse
Grundlage der Studie ist zunächst eine Literaturrecherche, um bestehende Definitionen, Modelle und Studien zu IKK national und international zu erfassen und auf ihre Angemessenheit für den Grundschulkontext hin zu diskutieren. Diese werden im Rahmen einer Dokumentenanalyse bildungspolitischen Forderungen auf nationaler Ebene gegenübergestellt. Die einleitende Klärung des Konstrukts IKK für den Englischunterricht im Primarbereich ist die Grundlage für die Entwicklung der Aufgaben.

Lehrwerksanalyse
Das Lehrwerk spielt im Englischunterricht eine zentrale Rolle, da sich viele Lehrer/innen bei der Planung, Strukturierung und Durchführung an den Angeboten des Kursmaterials orientieren. Der Stellenwert, den IKK in den Units der Lehrwerke einnimmt, wird sich somit auch implizit im Unterricht einer Lehrperson widerspiegeln. Exemplarisch wurden sechs Units aus drei Lehrwerken analysiert, um zu ermitteln, ob und wenn ja, wie die Lehrwerke IKK aufbereiten und über Themen und Aufgaben umsetzen.

Problemzentrierte Leitfadeninterviews
Grundlage der Interviews war ein speziell für diese Befragung konzipierter Interviewleitfaden, der zuvor in einer Pilotphase erprobt wurde. Jedes Interview wurde mit dem Einverständnis der Befragten aufgezeichnet. „Dies ermöglicht zum einen dem Interviewer, sich ganz auf das Gespräch einzulassen, ohne ständig Notizen machen zu müssen, und zum anderen wird im Nachhinein für eine unverfälschte Wiedergabe der gesprochenen Interviews gesorgt“ (Kuckartz et al.

2008: 26). Die transkribierten Interviews wurden inhaltsanalytisch ausgewertet. Die Qualitative Inhaltsanalyse (vgl. Mayring 2008) ermöglichte es, die subjektiven Perspektiven und Erfahrungen der Lehrpersonen zu rekonstruieren, da die Qualitative Inhaltsanalyse nicht nur eine bloße Technik, sondern immer an den Gegenstand und das Material der Analyse angebunden ist (vgl. ebd.: 44). Das regelgeleitete und systematische Vorgehen der Inhaltsanalyse, welches die Verwendung eines Kategoriensystems einschließt (vgl. ebd.: 42ff.), erlaubte es, die Vorstellungen der Lehrer/innen in Bezug auf IKK und deren Umsetzungen im Englischunterricht herauszuarbeiten.

3.2 Die Hauptstudie

Multiple Fallstudien

Ursprünglich sollten aus der Lehrerbefragung (Vorstudie) unterschiedliche Lehrertypen generiert werden, um daraus geeignete Lehrer/innen für die empirische Hauptstudie gewinnen zu können. Zum einen hat sich jedoch das erhobene Datenmaterial für eine Klassifizierung von Lehrertypen als nicht geeignet erwiesen. Zum anderen erklärten sich nur sehr wenige Lehrer/innen bereit, an der empirischen Hauptstudie teilzunehmen. Allerdings zeigte besonders eine Lehrperson ein starkes Interesse an der Teilnahme an der empirischen Hauptstudie. Es gelang ihr, noch weitere Lehrer/innen an der eigenen Grundschule für die Hauptstudie zu gewinnen. Insgesamt konnten drei Lehrer/innen für die empirische Studie an der Schule gewonnen werden. Eine/r der drei Lehrer/innen bot zusätzlich an, die empirische Studie in zwei von ihren Englischklassen durchführen zu können. So standen insgesamt vier Englischklassen für die Hauptstudie zur Verfügung. „Es wird immer wieder unterstrichen, dass der Begriff ‚Fall' nicht ausschließlich auf individuelle Personen und ihre Lebens-, Erziehungs- oder Bildungsgeschichte – bzw. ihre Problem- oder Krankengeschichte – beschränkt werden dürfe, sondern auch auf Gruppen (z.B. eine Familie, eine Schulklasse, eine Jugendclique) (…) angewendet werden könne" (Fatke 2010: 164f.). Da es sich um vier Schulklassen, also vier ‚Fälle' handelt, können diese nach Drese (2007: 193) als „multiple Fallstudien" definiert werden. Die Begleitung der vier Schulklassen an einer Grundschule erleichtert nicht nur die Durchführbarkeit der Studie. Von Vorteil ist auch, dass die Rahmenbedingung ‚gesamter Schulkontext' eine Konstante ist und so ein besonderer Fokus auf die vier individuellen Klassen und die Begleitung der drei unterschiedlichen Lehrer/innen gelegt werden konnte. Ziel dieses Vorgehens ist es, anhand von vier konkreten Fällen zu verstehen, welche Aspekte der Aufgaben Lehrer/innen bei der Planung eines interkulturellen kommunikativen Englischunterrichts diskutieren

und wie die Lehrer/innen das Potenzial der neu entwickelten Aufgaben nach der Durchführung im Unterricht im Hinblick auf IKK beurteilen.

Participatory Action Research
Participatory Action Research (PAR) lässt sich der Handlungsforschung zuordnen, welche in Publikationen und Studien neueren Datums verstärkt Beachtung gefunden hat (Allwright/Hanks 2009, Heigham/Croker 2009, Yoshida et al. 2009, Burns 2010, Friebertshäuser et al. 2010). PAR ist eine Methode, die nicht ‚über' Lehrer/innen aus einer distanzierten Perspektive forscht, sondern vielmehr kooperativ mit ihnen neues Wissen generieren möchte (vgl. McIntyre 2008: xii). So kann die Perspektive der Lehrer/innen verstanden und Implikationen für die Gestaltung zukünftiger Unterrichtsprozesse abgeleitet werden. Im Zentrum steht das gemeinsame Interesse von Teilnehmer/innen und Forscher/in, die Forschungsfrage unter den gegebenen Rahmenbedingungen eines spezifischen Feldes gemeinsam zu beantworten. Das Bündnis zwischen Lehrer/innen und Forscher/in bildet bei PAR die Grundlage für einen gelingenden Forschungsprozess (vgl. ebd.: 1). Dabei werden „Forschung und Entwicklung […] nicht methodologisch voneinander getrennt, sondern als aufeinander zu beziehende Teile ein und desselben Vorhabens verstanden" (Altrichter et al. 2010: 803). Den Forschungsprozess charakterisiert ein zyklisches Vorgehen. Die Phasen *Planning, Action, Observation* und *Reflection* können zyklisch und/oder iterativ verlaufen, bis ein zufriedenstellendes Ergebnis erreicht ist.(vgl. Burns 2010: 7). Dieses Vorgehen erlaubt es, die entwickelten Aufgaben gemeinsam mit den Lehrer/innen auf den jeweiligen Kontext anzupassen, den aus den Aufgaben resultierenden Prozess effektiv zu begleiten und abschließend auszuwerten. Durch eine Perspektiventriangulation können zudem Verzerrungen durch die Person des Forschers/der Forscherin aufgedeckt bzw. minimiert werden (vgl. Flick 2011: 14).

Die Datenerhebung
Die empirische Hauptstudie bedient sich unterschiedlicher Erhebungsmethoden. Dieses Vorgehen wird als Methodentriangulation (*between-method*) (vgl. Flick 2011: 15f.) bezeichnet. Die Kombination verschiedener Methoden stellt „weniger ein Instrument der Überprüfung empirischer Ergebnisse als einen Weg zu erweiterten Erkenntnismöglichkeiten [dar]" (ebd.: 9). Dabei ist zu beachten, „dass jede Methode ihren Gegenstand konstituiert (…) [und somit] unterschiedliche Konstruktionen eines Phänomens [zeigt]" (ebd.: 25). Der Vorteil einer Kombination von Methoden ist also der zusätzliche Erkenntnisgewinn im Hinblick auf den Untersuchungsgegenstand. Das gleiche Ziel verfolgt die Triangulation

von Perspektiven (vgl. ebd.: 20ff.). In dieser Studie werden die Annahmen, Perspektiven und Wahrnehmungen der Lehrer/innen und Schüler/innen eingeholt und mit denen der Forscher/in trianguliert. So können emische Sichtweisen (Innenperspektive der Forschungspartner (Lehrer/innen und Schüler/innen) mit der etischen Perspektive (Außenperspektive der Forscher/in) abgeglichen werden (vgl. Drese 2007: 192).

Im Folgenden werden die unterschiedlichen Erhebungsmethoden und Perspektiven vorgestellt, ihre unterschiedlichen Erkenntnisinteressen herausgearbeitet und aufgezeigt, wie durch ihre Kombination ein Erkenntniszugewinn möglich ist.

Die Unterrichtsdokumentation: Videografie

Bei der Durchführung der Aufgaben wurden alle Unterrichtsstunden mit dem Einverständnis der Beteiligten videografiert. Die Unterrichtsmitschnitte dürfen jedoch nur in anonymisierter Textform veröffentlicht werden. Mithilfe dieses Vorgehens konnte die Komplexität des Unterrichtsgeschehens (z.B. Interaktionen, Entscheidungen der Lehrer/innen, non-verbales Verhalten, Reaktionen und Bearbeitungsprozesse der Schüler/innen u.v.m.) festgehalten werden. „Ihr [der Videografie] wissenschaftlicher und forschungsmethodischer Mehrwert liegt auch in der relativ theorieunabhängigen und komplexitätserhaltenden Dokumentation, der theoretisch unbegrenzt häufigen Reproduzierbarkeit zu Analysezwecken" (Schramm/Aguado 2010: 186). Um das komplexe Geschehen im Klassenzimmer möglichst differenziert erfassen zu können, wurde der Empfehlung von Walsh (2011: 68) gefolgt, zwei Kameras einzusetzen. Auch wenn die Filmaufnahmen eine sehr alltagsnahe Erhebungssituation darstellen, so sind die Kameras doch ein Fremdkörper im Klassenzimmer, der die Faszination vieler Kinder auf sich zieht. Um diese Reaktivität zu reduzieren, lernten die Kinder zu Beginn der Aufnahmen die Kameras in einer Probestunde kennen, d.h. jede/r Schüler/in durfte sich die Kameras genau betrachten, Fragen stellen und durch das Objektiv schauen. Das Vertrautmachen mit der Technik führte dazu, dass die Mehrheit der Kinder das Interesse an den Kameras verlor.

Die Videoaufnahmen des Unterrichts wurden durch Transkription in Texte (vgl. Flick 2010: 314) umgewandelt und bildeten so die Grundlage für weitere Analysen. Hier wurde ein Nachteil der Methode deutlich. Dadurch, dass viele Schüler/innen gleichzeitig in den Gruppenarbeitsphasen mit ihren Mitschüler/innen sprachen, konnten diese individuellen Unterhaltungen nicht von den Kameras erfasst werden und hinterher somit auch nicht verschriftlicht werden. Gleiches gilt für Unterrichtssituationen, in denen mehrere Schüler/innen sich

zeitgleich äußerten oder es laute Nebengeräusche im Klassenzimmer gab. Da dieser Umstand bereits während der Videoaufnahmen auffiel, wurden die Kinder in der nächsten geplanten Gruppenarbeit mit jeweils einem Audiogerät versorgt. In der Unterrichtsstunde selber zeigte sich dann jedoch, dass zum einen das Verteilen der Geräte den Unterricht unterbrach und zum anderen die Schüler/innen durch die Geräte sehr abgelenkt waren. So wurde auf den zusätzlichen Einsatz der Audiogeräte verzichtet und dieser Nachteil der Videoaufnahmen in Kauf genommen. Die Videoaufnahmen an sich müssen jedoch bereits als Interpretation verstanden werden, da sie nur einen selektiven Ausschnitt der Wirklichkeit aufzeichnen.

Teilnehmende Beobachtung

In Ergänzung zu den Videoaufnahmen nahm der/die Forscher/in an jeder Unterrichtsstunde teil. Im Sinne der offenen Beobachtung (vgl. Lamnek 2005: 560f.) wurde den Lehrer/innen und Schüler/innen die Rolle des Forschers/der Forscherin dargelegt. Genau wie die Videografie erlaubt die teilnehmende Beobachtung das Verhalten der Teilnehmer/innen aus einer Außenperspektive zu beobachten (vgl. Cowie 2009: 166). Allerdings steht der Beobachter bei der teilnehmenden Beobachtung „nicht passiv-registrierend außerhalb seines Gegenstandsbereiches, sondern nimmt selbst teil an der sozialen Situation, in der der Gegenstand eingebettet ist. (…) [E]r sammelt Daten, während er an deren natürlichen Lebenssituation partizipiert“ (Mayring 2002: 80). Obwohl so die gesamte Komplexität der Unterrichtsprozesse in ihrem natürlichen Umfeld beobachtet werden konnte, entsteht bei jeder Beobachtung das Phänomen des *observer's paradox* (Nunan 2013: 595). Dieses besagt, dass ein Paradoxon entsteht, wenn man beobachten möchte, wie sich Menschen ohne systematische Beobachtung in ihrer natürlichen Umgebung verhalten, diese dafür aber systematisch beobachten muss (vgl. ebd.). Jede/r teilnehmende beobachtende Forscher/in muss sich dessen bewusst sein.

Während der Unterrichtsstunden wurden Aspekte notiert, die dem/der Forscher/in aus seiner/ihrer Außenperspektive relevant erschienen und mit den Lehrer/innen besprochen werden sollten. In Ergänzung zu dieser unstrukturierten Beobachtung (vgl. Lamnek 2005: 559f.) erlaubten die vorab entwickelten Aufgabendeskriptoren die Beobachtungen zu systematisieren und strukturieren. Dadurch, dass der/die Forscher/in durch die teilnehmende Beobachtung ein Teil der Unterrichtssituation geworden ist, entsteht somit automatisch eine Reaktivität mit dem Feld. In der qualitativen Forschung ist es absolut notwendig, dass der/die Forscher/in seine/ihre Rolle reflektiert (vgl. Bortz/Döring 2006: 339).

Sowohl die Videoaufnahmen als auch die Unterrichtsbeobachtungen bilden das Unterrichtsgeschehen von außen ab (Außenperspektive). Will man jedoch verstehen, welche Bedeutungen die Teilnehmenden mit ihren Handlungen verbinden, bedarf es auch der Erforschung, wie Lehrer/innen und Schüler/innen den Unterricht wahrnehmen (Innenperspektive). Deshalb wurden lehrer- und schülerbezogene Datensätze in Form von Beobachtungsbögen, Interviews, Fragebögen und Unterrichtserzeugnissen mit in die Studie einbezogen.

Lehrerbezogene Datensätze

Den Lehrer/innen kommt in dieser Studie eine besondere Rolle zu. Sie sind Partner/innen im Forschungsprozess (PAR) und unterstützen diesen mit ihrer kontextspezifischen Expertise sowie ihrem Erfahrungswissen bezogen auf ihren Englischunterricht und ihre Schüler/innen. Das Wissen über ihren Kontext brachten die Lehrer/innen in verschiedenen Gesprächen mit ein und hielten ihre Eindrücke des Unterrichtsgeschehens in Beobachtungsbögen fest.

Vorbesprechung 1 mit den Lehrer/innen (Gruppengespräch)

Das erste Gespräch fand in Form eines gemeinsamen Gruppentreffens an der Schule der Lehrer/innen statt. Zunächst wurden die Lehrer/innen in das Forschungsvorhaben eingeführt und das Ziel der Studie erläutert. „[P]ractitioners of PAR have an obligation at the beginning of a PAR project to let their desires be known to research participants about how they, as researchers, intend to analyze and disseminate research data“ (McIntyre 2008: 54). Weiterhin wurden die Rollen im Forschungsprozess geklärt. Der/Die Forscher/in plante und begleitete die empirische Studie. Dazu zählten auch die theoretisch fundierte Planung des Unterrichts und die Entwicklung der Materialien. Die Lehrer/innen hatten die Aufgabe die Unterrichtsplanungen und Materialien vorher immer auf ihre Klasse und die Machbarkeit zu prüfen. Diesbezüglich sollten sie dem/der Forscher/in Rückmeldung geben, damit die Planungen und Materialien bei Bedarf überarbeitet werden konnten. Der Unterricht an sich wurde von den Lehrer/innen durchgeführt. Der/Die jeweilige Lehrer/in blieb somit die gewohnte Bezugsperson für die Schüler/innen.

Vorbesprechung 2 mit den Lehrer/innen (Einzelgespräch)

Ziel der Einzelgespräche war es, individuell mit jedem/r Lehrer/in die Entwürfe der Unterrichtsplanungen und Materialien in Bezug auf die jeweilige Klasse abzustimmen. Insgesamt wurden die Planungen sehr offen und positiv

aufgenommen und es mussten nur kleinere Überarbeitungen (wie z.B. Begrüßungsformen, Rituale) vorgenommen werden.

Gespräche und Absprachen im Erhebungszeitraum
Immer am Ende einer Unterrichtswoche erhielten die Lehrer/innen die beiden Verlaufspläne mit allen Materialien für die darauffolgende Woche. So hatten die Lehrer/innen stets genug Zeit, sich auf die neue Woche vorzubereiten, Unklarheiten mit dem/der Forscher/in abzusprechen und die ausgearbeiteten Unterrichtsideen in Bezug auf Machbarkeit und Relevanz für ihren Kontext zu überprüfen.

Nachbesprechung mit den Lehrer/innen
Nach Abschluss der Unterrichtseinheit in jeder Klasse wurde mit jedem/r Lehrer/in ein umfassendes, retrospektives Einzelgespräch (ca. 60 Minuten) geführt, welches mit einem Audiogerät aufgezeichnet wurde. Grundlage der Interviews war ein Interviewleitfaden. Die Fragen waren so formuliert, dass sie die Lehrer/innen zum Erzählen anregten, denn es war von Interesse, zu erfahren, wie die Lehrer/innen die Unterrichtseinheit wahrgenommen hatten und wie sie aus ihrer Sicht das Potenzial der Aufgaben und Materialien im Hinblick auf die Entwicklung von IKK bewerten.

Beobachtungsbögen
Das Abschlussgespräch bezog sich auf die Reflexion der gesamten Unterrichtseinheit. Im Rahmen dieser Interviews wäre es aber nicht zielführend gewesen, alle Unterrichtsstunden einzeln im Hinblick auf das Potenzial der Aufgaben zu besprechen. Dies wäre zum einen zu umfangreich geworden und zum anderen wären bis zum Ende der Einheit manche Erinnerungen eingetrübt oder durch andere Ereignisse überlagert worden. Um aber eine Rückmeldung zu den Aufgaben in jeder einzelnen Englischstunde zu erhalten, sind Beobachtungsbögen ein geeignetes Instrument, da sie eine „systematische Beobachtung" (Hochstetter 2011: 39) des Unterrichtsgeschehens erlauben.

Die Erstellung der verwendeten Beobachtungsbögen für die empirische Studie unterlag einem Entwicklungsprozess. Bei dem ersten Treffen mit den Lehrer/innen (Gruppengespräch) wurde die Bereitschaft der Lehrer/innen erfragt, einen Beobachtungsbogen (Umfang zwischen 1 und 4 Seiten) nach jeder Stunde auszufüllen, da dieses mit einem zusätzlichen Zeitaufwand für die Lehrer/innen verbunden war. Alle zeigten sich jedoch einverstanden und offen, denn über das konkrete Schülerverhalten (*task-in-process*) konnte geschlussfolgert

werden, ob die Aufgabenkriterien das Potenzial hatten, IKK anzubahnen. Dennoch sind „für die interkulturelle Kompetenz gewünschte Haltungen und Einstellungen zur Zielsprache, zur Kultur in einem Zielsprachenland oder allgemein zu Fremden besser über Befragungen, Diskussionen oder andere Methoden zu ermitteln als ausschließlich über Beobachtungen“ (ebd.: 42f.). Aus diesem Grund sind die Beobachtungsbögen nur eine Methode, um Einblicke in das Unterrichtsgeschehen zu erhalten. Die Stärke dieses Instruments liegt in der systematischen und somit eher objektiven Beurteilung des Unterrichts und ist zum anderen ein Format, mit dem die Lehrer/innen (z.B. aus Lehrwerken) vertraut sind.

Zusätzlich zu den Lehrer/innen trug auch der/die Forscher/in, zeitnah nach jeder Englischstunde, ihre Beobachtungen in den klassenspezifischen Beobachtungsbogen ein. So konnten zwei Perspektiven auf dasselbe Geschehen gegenübergestellt werden. Gleiche Beobachtungen verstärkten die Annahme, dass die Aufgaben das Potenzial für die Anbahnung von IKK haben. Gegenläufige oder abweichende Wahrnehmungen hingegen schwächten diese Annahmen, waren jedoch interessante Ausgangspunkte für das abschließende Gespräch mit den Lehrer/innen.

Schülerbezogene Datensätze

Die Schüler/innen bestimmen zu einem wesentlichen Teil, wie sich der *task-as-workplan* im Unterrichtsgeschehen entfaltet. In der Studie können das Schülerverhalten (Videografie) und Unterrichtserzeugnisse (wie z.B. Arbeitsblätter, Lernertexte) der Kinder ebenfalls genutzt werden, um das interkulturell kommunikative Potenzial von Aufgaben einzuschätzen. Darüber hinaus wurden Fragebögen (Klassen 3 und 4) eingesetzt und Gruppeninterviews (in allen Klassenstufen) geführt, um zu erfahren, wie die Schüler/innen den Unterricht wahrgenommen haben.

Unterrichtserzeugnisse

Während der Unterrichtseinheit entstanden in allen vier Klassen unterschiedliche Schülerprodukte (z.B. Arbeitsblätter, E-Mails, Poster, Geburtstagskalender). In den beiden ersten Klassen enthalten diese primär gestalterische Elemente (wie z.B. der Geburtstagskalender). In den Klassen 3 und 4 geben über die gestalterischen Aspekte hinaus schriftliche Erzeugnisse (beispielweise E-Mails der Schüler/innen) Einsichten in interkulturell kommunikative Lernprozesse. Über die Schülerprodukte konnten „Einblicke in Kinderwelten“ (Mey 2003: 5) gewonnen werden.

Fragebögen
Am Ende der Unterrichtseinheit wurden die Schüler/innen in den Klassen 3 und 4 gebeten, einen Fragebogen auszufüllen. In den beiden ersten Klassen wurde auf den Einsatz des Fragebogens verzichtet, da das Erlesen der Fragen und der Einsatz der Schrift die jungen Schüler/innen überfordert hätten (vgl. ebd. 2003: 5f.). Die Vorteile von diesem Vorgehen sind, dass in einer sehr geringen Zeit Informationen von vielen Personen gleichzeitig zu ein und demselben Gegenstand erfragt werden können (vgl. Dörnyei 2010: 6). So konnte ein allgemeiner Überblick über die Meinungen, Perspektiven und die Stimmungslage der Schüler/innen in beiden Klassenstufen gewonnen werden.

Bei der Entwicklung des Fragebogens wurde auf eine übersichtliche Gestaltung und die richtige Lineatur für die jeweilige Klassenstufe geachtet. Insgesamt umfasste der Fragebogen 14 Fragen, welche die Schüler/innen auf 4 Seiten beantworten sollten. Dies entspricht Dörnyeis (vgl. ebd.: 12f.) Vorgaben, denn ein Fragebogen sollte eine Bearbeitungszeit von 30 Minuten und den Umfang von 4 bis 6 Seiten nicht überschreiten, da bei dieser jungen Zielgruppe die Schreib- und Lesekompetenzen noch nicht ausgereift sind und auch die Konzentrationsfähigkeit schnell erschöpft ist. Weiterhin ist es wichtig, eine für die Kinder verständliche und angemessene Sprache zu verwenden, d.h., dass die Fragen klar, konkret und einfach formuliert sein müssen (vgl. Mey 2003: 7). Die Fragen sprachen die Schüler/innen immer direkt an und fragten sie nach ihrer subjektiven Meinung. Zum einen zeigte dies den Kindern, dass ihre individuelle Meinung wichtig ist und zum anderen wurde für die Schüler/innen so auch deutlich, dass es eine ‚falsche' Antwort nicht gibt und es um ihre Wahrnehmung der Dinge geht. Die ersten Fragen erfragten soziodemographische Daten der Schüler/innen und erhoben ihre Meinung zu ihrem Englischunterricht. Durch den persönlichen Bezug und die Allgemeinheit der Fragen merkten die Kinder, dass sie die Fragen beantworten konnten und gewannen so zu Beginn Sicherheit. Diese stärkte sie für den weiteren Verlauf der Fragen (vgl. Dörnyei 2007: 111f.), da diese zunehmend spezifischer wurden.

Der Fragebogen bestand überwiegend aus offenen Fragen, weil die Antworten auf diese mehr Detailinformationen enthalten und eventuell sogar Aspekte (*issues*) aufzeigen, die der/die Forscher/in nicht antizipiert hat. „[O]pen-format items can provide a greater ‚richness' (…) [and] yield graphic examples, illustrative quotes and can also lead us to identify issues not previously anticipated" (Dörnyei 2010: 26). Die geschlossenen Fragen enthielten nur die beiden sehr eng gefassten Antwortmöglichkeiten ‚Ja' und ‚Nein', so dass die Schüler/innen hierbei eine Tendenz ausdrücken mussten. Es bestand jedoch immer die Möglichkeit (mit einer Ausnahme), die ausgewählte

Option zu begründen, um so die Hintergründe für die angekreuzte Antwort zu erfassen. Auch wenn der Fragebogen den Schüler/innen die Möglichkeit bot, ihre eigenen Antworten einzubringen, war dies natürlich nur begrenzt möglich. Diese Begrenzung liegt in der Natur der Fragebögen an sich. Um mehr in die Tiefe gehen zu können, wurden zusätzlich Gruppeninterviews geführt. „Um die mit Fragebögen verbundenen Einschränkungen und Herausforderungen zu kompensieren, wird vielfach außerdem empfohlen, Kinder zusätzlich zu interviewen, um für bestimmte Fragebereiche ergänzendes Datenmaterial zu erhalten, das für die Interpretation kindlicher Perspektiven nützlich ist“ (ebd.: 8).

Gruppeninterviews
Während die Fragebögen ein Meinungsbild aller Dritt- und Viertklässler erheben, sind Interviews besonders geeignet, um einzelne Aspekte differenzierter zu thematisieren. Wie auch bei den Gesprächen mit den Lehrer/innen sind es erneut die subjektiven Wahrnehmungen, diesmal die der Kinder, die im Mittelpunkt des Interesses stehen. „Interviews [eröffnen] (…) einen wichtigen Zugang zu kindlichem Wissen“ (Mey 2003: 8). Die Interviews wurden in allen Klassenstufen mit einer Auswahl an Schüler/innen geführt.

Ein leitfadengestütztes Interview strukturierte die Gespräche. Dies ist ein wichtiges Kriterium, denn Kinder im Grundschulalter entwickeln erst die Fähigkeit, ausführlich und detailliert über Sachverhalte zu berichten. Ein narratives Interview hätte die Kinder somit überfordert (vgl. ebd.: 9). Dies besagt zugleich auch, dass sich die Sprechzeit des interviewenden Erwachsenen in einem Gespräch mit Kindern erhöht. Dies entspricht jedoch eigentlich nicht dem Grundgedanken eines leitfadengestützten Interviews (vgl. Mayring 2002: 66f.). Um die Sprechzeit der Kinder zu erhöhen, wurde deshalb ein Gruppeninterview gewählt. „Gruppendiskussionen [erlauben] Einblicke in die jeweiligen Kinderkulturen und eine Beschreibung von Kinderwelten ‚von innen‘ her“ (Mey 2003: 13). Kinder regen sich in Gruppen gegenseitig zum Sprechen an und es entstehen Diskussionen, die gewünscht sind. Aus diesem Grund kann man eher von einem Gruppengespräch sprechen, dessen Inhalte und Richtung aber durch einen Leitfaden bestimmt waren. Dabei ist die Zusammensetzung der Gruppen wichtig. Es ist „empfehlenswert, Gruppendiskussionen in ‚Realgruppen‘ durchzuführen, also in Gruppen, die unabhängig von der Untersuchung existieren (statt eine Gruppe zu Forschungszwecken kurzfristig und einmalig zusammenzustellen), um Kindern eine ihnen vertraute Atmosphäre zu bieten“ (ebd.: 13). So kamen die Schüler/innen immer mit ein bis zwei vertrauten

Klassenkamerad/innen zum Gespräch. Die Gruppengröße aller Interviews lag zwischen zwei und drei Kindern. In jeder Klasse stand die letzte Schulstunde der Unterrichtseinheit zur Verfügung, in der zwei Gruppen à 20 Minuten interviewt werden konnten. Alle Schülerinterviews wurden mit einer Videokamera aufgezeichnet, um zusätzlich zu den verbalen Aussagen, non-verbale Reaktionen mit erfassen zu können.

„Auf *inhaltlicher Ebene* kann u.a. beobachtet werden, welche Erfahrungen und welches Wissen Kinder zu einem Thema beitragen und mit welchen Argumenten sie ihre Ausführungen stützen. Das gibt auch einen Einblick in die Orientierungsmuster, mit denen sie ihre Alltagserfahrungen selbst einordnen und deuten" (Billmann-Mahecha 2001: 14). Dabei müssen die Fragen, wie bei den Fragebögen, im Gespräch einen Bezug zu der direkten Lebenswelt aufweisen und klar formuliert sein. Aus diesem Grund wurden zwei unterschiedliche Leitfäden für die ersten beiden Klassen und die Klassen 3 und 4 konzipiert.

4. Zusammenfassung, Schlussfolgerungen und Ausblick

Die Forschungsarbeit hat das Ziel, theoretische und praktische Erkenntnisse in Bezug auf die Anbahnung von IKK im Englischunterricht der Grundschule hervorzubringen. Da in diesem Feld bisher nur sehr wenige empirisch überprüfte Erkenntnisse vorliegen, wurde ein qualitatives Forschungsdesign gewählt, welches sich der Thematik explorativ-interpretativ nähert. Anhand von vier Fallstudien sollen Möglichkeiten und Grenzen der Vermittlung von IKK im Grundschulkontext aufgezeigt werden. Die Forschungsarbeit erhebt jedoch nicht den Anspruch, diese Erkenntnisse als wissenschaftliche Hypothesen zu präsentieren. Denn dadurch, dass sich die Studie auf vier konkrete Fälle sowie deren Rahmenbedingungen, Akteure und Handlungsverläufe bezieht, gelten die gewonnenen Erkenntnisse speziell für diese und sind eigentlich nicht generalisierbar (vgl. Schocker-von Ditfurth 2001: 106). Allerdings kann „aus dem Besonderen des Einzelfalls durchaus Allgemeines […] entstehen" (Fatke 2010: 168). „Der Geltungsbereich einer Studie und damit auch der Grad ihrer Generalisierbarkeit wird deshalb zunächst von der Vergleichbarkeit der Kontexte abhängen, für die die Ergebnisse der Studie herangezogen werden" (Drese 2007: 214). Die Rezipienten (Forscher/innen und Lehrer/innen) dieser Forschungsarbeit können dann individuell prüfen, inwiefern die Ergebnisse der Forschungsarbeit auch für den Kontext, der sie betrifft bzw. in dem sie tätig sind, relevant sind.

Literatur

Allwright, Dick / Hanks, Judith (2009). *The Developing Language Learner. An Introduction to Exploratory Practice.* Basingstoke: Palgrave Macmillan.

Altrichter, Herbert / Aichner, Waltraud / Soukup-Altrichter, Katharina / Welte, Heike (2010). PraktikerInnen als ForscherInnen. Forschung und Entwicklung durch Aktionsforschung. In: Friebertshäuser, Barbara / Langer, Antje / Prengel, Annedore (Hrsg.). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft.* 3. Aufl. Weinheim, München: Juventa, 803-818.

Bennett, Milton J. (1986). A Developmental Approach to Training for Intercultural Sensitivity. *International Journal of Intercultural Relations* 10/2, 179-196.

Billmann-Mahecha, Elfriede (2001). Soziale Aushandlungsprozesse im Kindesalter – ein qualitativer Zugang über das Gruppendiskussionsverfahren. In: Mey, Günter (Hrsg.). *Qualitative Forschung in der Entwicklungspsychologie. Potentiale, Probleme, Perspektiven.* Forschungsbericht aus der Abteilung Psychologie im Institut für Sozialwissenschaften der Technischen Universität Berlin. 2001/1, 12-18.

Bortz, Jürgen / Döring, Nicola (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler.* 4. Aufl. Berlin: Springer.

Brusch, Wilfried (2000). Englisch: die vierte Kulturtechnik? *Grundschule* 32/6, 34.

Burns, Anne (2010). *Doing Action Research in English Language Teaching. A Guide for Practitioners*. New York: Routledge.

Byram, Michael (1997). *Teaching and Assessing Intercultural Competence*. Clevedon: Multilingual Matters.

Byram, Michael / Gribkova, Bella / Starkey, Hugh (2002). *Developing the Intercultural Dimension in Language Teaching. A Practical Introduction for Teachers*. Strasbourg, Graz: Council of Europe / ECML.

Cameron, Lynne (2001). *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press.

Caspari, Daniela / Helbig, Beate / Schmelter, Lars (2003). Forschungsmethoden: Exploratives-Interpretatives Forschen. In: Bausch, Karl-Richard / Christ, Herbert / Krumm, Hans-Jürgen (Hrsg.). *Handbuch Fremdsprachenunterricht*. 5. Aufl. Tübingen: Francke, 499-505.

Christ, Herbert (2003). Erwerb von Fremdsprachen im Vorschul- und Primarbereich. In: Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.). *Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 449-454.

Christ, Herbert (2007). Fremdverstehen in der Praxis interkulturellen Lernens im Fremdsprachenunterricht. In: Bredella, Lothar / Christ, Herbert (Hrsg.). *Fremdverstehen und interkulturelle Kompetenz.* Tübingen: Narr, 51-77.

Council of Europe (Hrsg.) (2001). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment.* Cambridge: Cambridge University Press.

Cowie, Neil (2009). Observation. In: Heigham, Juanita / Croker, Robert A. (Hrsg.). *Qualitative Research in Applied Linguistics. A Pratical Introduction.* Basingstoke: Palgrave Macmillan, 165-181.

Croker, Robert A. (2009). An Introduction to Qualitative Research. In: Heigham, Juanita / Croker, Robert A. (Hrsg.). *Qualitative Research in Applied Linguistics. A Practical Introduction.* Basingstoke: Palgrave Macmillan, 3-24.

Dörnyei, Zoltán (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Dörnyei, Zoltán (2010). *Questionnaires in Second Language Learning Research. Construction, Administration, and Processing.* 2. Aufl. New York: Routledge.

Doyé, Peter (1999). *The Intercultural Dimension. Foreign Language Education in the Primary School.* Berlin: Cornelsen.

Drese, Karin (2007). *Einschätzung der Sprechleistung von Lernern im Englischunterricht der Grundschule*. Dissertation. Gießen: Universität Gießen.

Fatke, Reinhard (2010). Fallstudien. In: Friebertshäuser, Barbara / Langer, Antje / Prengel, Annedore (Hrsg.). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. 3. Aufl. Weinheim, München: Juventa, 159-172.

Flick, Uwe (2011). *Triangulation. Eine Einführung.* 3. Aufl. Wiesbaden: Springer.

Flick, Uwe (2010). *Qualitative Sozialforschung. Eine Einführung.* 3. Aufl. Reinbek bei Hamburg: Rowohlt.

Freitag-Hild, Britta (2010). *British Fictions of Migration im Fremdsprachenunterricht – Theorie, Aufgabentypologie und Unterrichtspraxis inter- und transkultureller Literaturdidaktik*. Trier: wvt.

Friebertshäuser, Barbara / Langer, Anja / Prengel, Annedore (Hrsg.) (2010). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft.* 3. Aufl. Weinheim, München: Juventa.

Guilherme, Manuela (2013). Intercultural Competence. In: Byram, Michael / Hu, Adelheid (Hrsg.). *Routledge Encyclopedia of Language Teaching and Learning*. 2. Aufl. Oxon: Routledge, 346-349.

Hammer, Mitch / Bennett, Milton J. (2001). The Developmental Model of Intercultural Sensitivity. In: *Intercultural Development Inventory Manual*, 12-16.

Heigham, Juanita / Croker, Robert A. (Hrsg.) (2009). *Qualitative Research in Applied Linguistics. A Practical Introduction.* Basingstoke: Palgrave Macmillan.

Hessisches Kultusministerium (2010). *Bildungsstandards und Inhaltsfelder - Das neue Kerncurriculum für Hessen. Primarstufe. Moderne Fremdsprachen. (Entwurf).* Wiesbaden. Online: http://www.iq.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/IQ_Internet/med/d9d/d9d1d584-b546-821f-012f-31e2389e4818,22222222-2222-2222-2222-222222222222 [10.3.2014].

Hochstetter, Johanna (2011). *Diagnostische Kompetenz im Englischunterricht der Grundschule.* Tübingen: Narr.

Hu, Adelheid (2010). Interkulturelle Kommunikative Kompetenz. In: Hallet, Wolfgang / Königs, Frank G. (Hrsg.). *Handbuch Fremdsprachendidaktik.* Seelze-Velber: Klett / Kallmeyer, 75-79.

Hu, Adelheid / Byram, Michael (2009). Introduction. In: Hu, Adelheid / Byram, Michael (Hrsg.). *Interkulturelle Kompetenz und fremdsprachliches Lernen. Modelle, Empirie, Evaluation.* Tübingen: Narr, VII-XXV.

Institut für Qualitätsentwicklung (2011). *Leitfaden. Maßgebliche Orientierungstexte zum Kerncurriculum. Primarstufe. Wiesbaden* (im Auftrag des Hessischen Kultusministeriums). Online: http://www.iq.hessen.de/irj/IQ_Internet?rid=HKM_15/IQ_Internet/nav/232/2326c76d-1dbb-f21f-012f-31e2389e4818,,,,11111111-2222-3333-4444-100000005002%26_ic_seluCon=2441afe0-0a3e-f21f-012f-31e2389e4818%26shownav=false.htm&uid=2326c76d-1dbb-f21f-012f-31e2389e4818&shownav=false [10.3.2014].

Kuckartz, Udo / Dresing, Thorsten / Rädiker, Stefan / Stefer, Claus F. (2008). *Qualitative Evaluation – Der Einstieg in die Praxis.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Lamnek, Sigfried (2005). *Qualitative Sozialforschung. Lehrbuch.* 4. Aufl. Basel: Beltz.

Legutke, Michael K. / Müller-Hartmann, Andreas / Schocker-v. Ditfurth, Marita (2009). *Teaching English in the Primary School.* Stuttgart: Klett.

Legutke, Michael K. / Müller-Hartmann, Andreas / Schocker-von Ditfurth, Marita (2010). Frühbeginnender Fremdsprachenunterricht. In: Hallet, Wolfgang / Königs, Frank G. (Hrsg.). *Handbuch Fremdsprachendidaktik.* Seelze-Velber: Klett/Kallmeyer, 290-294.

Marschollek, Andreas (2002). *Kognitive und affektive Flexibilität durch fremde Sprachen. Eine empirische Untersuchung in der Primarstufe.* Münster: LIT.

Marschollek, Andreas (2011). Fremdsprachliches Lernen. In: Einsiedler, Wolfgang / Götz, Margarete / Hartinger, Andreas / Heinzel, Friederike / Kahlert, Joachim / Sandfuchs, Uwe (Hrsg.). *Handbuch Grundschulpädagogik und Grundschuldidaktik*. 3. Aufl. Regensburg: Klinkhardt, 455-461.

Mayring, Philipp (2002). *Einführung in die Qualitative Sozialforschung*. 5. Aufl. Weinheim und Basel: Beltz.

Mayring, Philipp (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 10. Aufl. Weinheim und Basel: Beltz.

McIntyre, Alice (2008). *Participatory Action Research.* Thousand Oaks: Sage Publications.

Mey, Günter (2003). *Zugänge zur kindlichen Perspektive. Methoden der Kindheitsforschung. Forschungsbericht aus der Abteilung Psychologie im Institut für Sozialwissenschaften der Technischen Universität Berlin.* 2003/1. Berlin: Technische Universität.

Ministerium für Kultus, Jugend und Sport Baden-Württemberg (2004). *Bildungsstandards für Englisch. Grundschule – Klassen 2,4.* Online: http://www.bildung-staerkt-menschen.de/service/downloads/Bildungsstandards/GS/GS_E_bs.pdf [04.07.2014].

Müller-Hartmann, Andreas / Schocker, Marita / Pant, Hans Anand (2013). *Kompetenzentwicklung in der Sek. I. Lernaufgaben Englisch aus der Praxis*. Braunschweig: Diesterweg.

Nunan, David (2013). Research methods. In: Byram, Michael / Hu, Adelheid (Hrsg.): *Routledge Encyclopedia of Language Teaching and Learning*. 2. Aufl. Oxon: Routledge. 593-598.

Pfitzer, Petra (2006). *Kindgemäßes Fremdsprachenlernen. Zur Theorie und Praxis des Frühbeginns Englisch.* Frankfurt/Main: Lang.

Schocker-von Ditfurth, Marita (2001). Die Suche nach einem gegenstandsangemessenen Ansatz zur Erforschung von Lernprozessen in komplexen pädagogischen Handlungsfeldern: Grundsätze und Verfahren ethnografischer Forschung. In: Müller-Hartmann, Andreas / Schocker-von Ditfurth, Marita (Hrsg.). *Qualitative Forschung im Bereich Fremdsprachen lehren und lernen*. Tübingen: Narr, 84-113.

Schocker-von Ditfurth, Marita (2004). Die Rolle des Englischen in der Diskussion um eine ‚Mehrsprachigkeitsdidaktik'. In: Bausch, Karl-Richard / Königs,

Frank G. / Krumm, Hans-Jürgen (Hrsg.). *Mehrsprachigkeit im Fokus. Arbeitspapiere der 24.Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen*: Narr, 215-225.

Schramm, Karen / Aguado, Karin (2010). Videografie in den Fremdsprachendidaktiken. Ein Überblick. In: Aguado, Karin / Schramm, Karen / Vollmer, Helmut Johannes (Hrsg.). *Fremdsprachliches Handeln beobachten, messen, evaluieren*. Frankfurt/Main: Lang, 185-214.

Walsh, Steve (2011). *Exploring Classroom Discourse. Language in Action.* Oxon: Routledge.

Yoshida, Tatsuhiro / Imai, Hiroyuki / Nakata, Yoshiyuki / Tajino, Akira / Takeuchi, Osamu / Tamai, Ken (Hrsg.) (2009). *Researching Language Teaching and Learning. An Integration of Practice and Theory.* Bern: Lang.

Methoden videogestützter Beobachtungsverfahren in der Lehr-Lern-Forschung

Johannes Appel / Udo Rauin

Mit wachsenden technischen Möglichkeiten erweist sich die Nutzung von Videoaufnahmen zunehmend als praktikable Erweiterung des Methodenrepertoires zur Erforschung von Unterricht. Beobachtungsstudien sind heute durch die Flexibilität des Datenmaterials weitaus vertiefter möglich, als es die klassische direkte Beobachtung im Klassenraum erlaubt. Neben Aspekten der Forschungspragmatik liegen die Vorteile besonders in Möglichkeiten der Kombination von Videodaten mit alternativ erhobenen Datenquellen, wie Befragungen oder Tests, um umfassendere Einblicke in das komplexe Geschehen im Unterricht zu ermöglichen. In diesem Beitrag werden neben einem Überblick zu Vor- und Nachteilen und methodischen Grundlagen des Zugangs zentrale Verfahren der Auswertung von audiovisuellen Unterrichtsaufzeichnungen dargestellt und Triangulationspotenziale – insbesondere der Datentriangulation – aufgezeigt. Zur Veranschaulichung werden Beispieldaten aus einem eigenen Forschungsprojekt herangezogen.

1. Videografie in der Unterrichtsforschung

Ihre Anfänge hat die videogestützte Unterrichtsforschung in den 1970er Jahren, als Technologien zum Anfertigen und Abspielen von Videoaufnahmen technisch einfach realisierbar und vergleichsweise kostengünstig wurden. Zu den Pionieren dieses Zugangs zählen beispielsweise Raymond S. Adams und Bruce J. Biddle, die bei ihrer Arbeit zum Interaktionsgeschehen zwischen Lehrern und Schülern die damals gängige direkte Beobachtung im Klassenraum als zu eingeschränkt und ungenau empfanden und daher mit Videoaufnahmen arbeiteten (Adams/Biddle 1970). Auch Jacob S. Kounin, der zur Untersuchung des *classroom managements* nach Möglichkeiten zur Erweiterung der Beobachtung im Klassenraum suchte und aus diesem Grund so genannte „Video-Recorder-Studien" durchführte (Kounin 1976), gilt als einer der Pioniere im Bereich der videogestützten Unterrichtsforschung. Die zentralen Vorteile, die man von Beginn an in der Verwendung von Videoaufnahmen sah, werden im Wesentlichen bis heute angeführt (vgl. Biddle 1967, Adams/Biddle 1970, Kounin 1976, Jacobs et al. 1999, Aufschnaiter/Welzel 2001, Petko et al. 2003, Dinkelaker/Herrle 2009, Pauli/Reusser 2006) und werden nachfolgend ausführlich dargelegt.

Ein großer Vorteil ergibt sich aus der Tatsache, dass Videodaten einen wiederholten Zugriff auf das audiovisuelle Geschehen der Klassensituation zulassen. Somit werden mehrere und voneinander unabhängige Analysen des Materials mit unterschiedlichen analytischen Zugängen und Fragestellungen ermöglicht. Dies erscheint zum einen unter methodologischen Gesichtspunkten interessant, weil auf diese Weise die Erprobung und der Vergleich von Beobachtungsinstrumentarien vereinfacht wird, und zum anderen Datenbestände aufgebaut werden können, die sich zur Nutzung in Forschungsverbünden oder für Sekundäranalysen anbieten. Erfüllen die Aufnahmen bestimmte Standards der Bild- und Tonqualität, besteht zudem ein weiterer Nutzen darin, dass die Daten durch Techniken wie Vor- und Zurückspulen, Zeitlupe oder Zoom tieferen Analysen zugänglich gemacht werden, als es dem menschlichen Beobachter unter normalen physikalischen Bedingungen möglich wäre. Das bezieht sich sowohl auf meist unbemerkte Ereignisse im Klassenraum (z.B. Briefchen unter dem Tisch, Gesten von Schülern), als auch auf die Gleichzeitigkeit verschiedener Handlungen in der Unterrichtssituation, die sich in ihrer Fülle und Komplexität der Aufmerksamkeit eines einzelnen Beobachters zwangsläufig entziehen. Videoaufnahmen ermöglichen somit zwar einen hohen Informationsgehalt und eine hohe Genauigkeit der Daten zu jenem Ausschnitt des Unterrichtsgeschehens, der in der Erhebungssituation fokussiert wird. Allerdings bleiben durch die unumgänglichen Selektionsentscheidungen im Rahmen der Aufzeichnungen notwendigerweise immer Aspekte des Geschehens unberücksichtigt, die mitunter – je nach Forschungsinteresse – auch durch teilnehmende Beobachtungen zugänglich gemacht werden können, wie beispielsweise Gesten, Äußerungen, Bewegungen außerhalb des Kamera- und Mikrofonbereichs oder schlicht Geruchs- und Temperaturwahrnehmungen in der jeweiligen Situation (Dinkelaker/Herrle 2009). Darüber hinaus besteht ein entscheidender Vorteil der Nutzung von Videodaten darin, dass durch die Trennung der apparativ basierten Datenerhebung und der zu einem späteren Zeitpunkt von den Forschenden vollzogenen Analyse eine klare Trennung zwischen Protokollierung und Interpretation der Ereignisse geschieht. Dadurch ist eine erhöhte Genauigkeit der Einschätzungen bei gleichzeitiger Reduzierung von Fehlentscheidungen möglich, weil beim Vorgang der Identifizierung und Bewertung der Ereignisse weder besonderer Zeitdruck besteht, noch kognitive Einschränkungen des Beobachters aufgrund von Ablenkungen oder „einem schlechten Tag" zu befürchten sind.

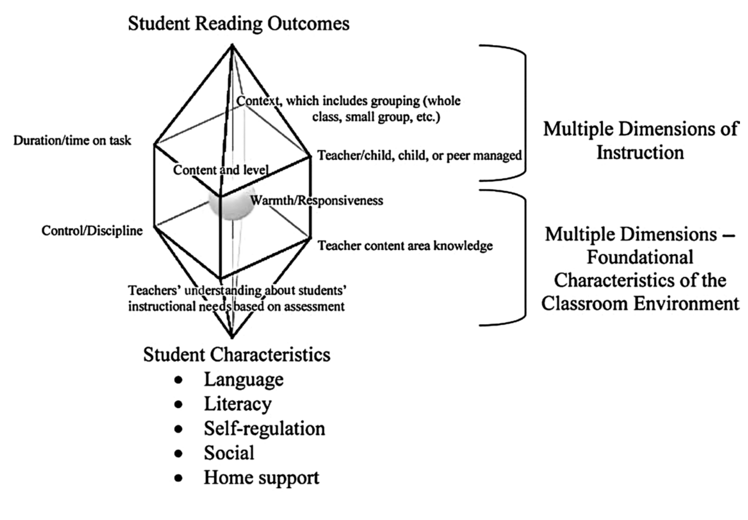
Neben diesen Vorteilen weiß man aufgrund der mittlerweile großen Anzahl von Videostudien der letzten Jahrzehnte auch um einige Herausforderungen, die man bei der Planung des Forschungsprozesses bedenken sollte (vgl. Seidel et al. 2003, Wild 2003, Dinkelaker/Herrle 2009). So ist bereits im Vorfeld die datenschutzrechtliche Absicherung zu beachten, zu der schriftliche Einverständniserklärungen

aller abzubildenden Personen bzw. ihrer Erziehungsberechtigten, inkl. der zuständigen Behörde (Schulleitung, Schulamt) einzuholen sind. Solche Vorgänge können durch bereits bestehende Kooperationen zu den Schulen bzw. Schulträgern unter Umständen erheblich erleichtert werden, sollten jedoch hinsichtlich des Aufwands und der Bedeutung nicht unterschätzt werden. Eine weitere Herausforderung besteht in der Realisierung eines tragfähigen Untersuchungsdesigns. So kann die Qualität der Stichprobe – mit Blick auf die jeweilige Fragestellung – durch organisatorische Einschränkungen der Datenerhebung beeinträchtigt werden (z.B. Verteilung von Personen- oder Gruppenmerkmalen in den zugänglichen Lerngruppen, gewünschte Klassenstufe oder Unterrichtsfach nicht verfügbar usw.). Nach Abschluss der Vorarbeiten erfordert die Anfertigung der Videoaufnahmen eine Reihe von Erwägungen, die sich aus dem Zusammenspiel technisch-organisatorischer und methodischer Gesichtspunkte ergeben. Denn möchte man eine möglichst umfassende Abbildung des Geschehens bei gleichzeitig hoher Qualität der Video- und Audioaufnahmen erzielen, ist der Einsatz entsprechend hochwertiger Geräte in ausreichender Anzahl notwendig. Beispielsweise sind für viele Untersuchungsgegenstände und Fragestellungen mindestens zwei Kameraperspektiven erforderlich, die im Falle variierender Raum- oder Gruppenanordnungen gegebenenfalls um zusätzliche Perspektiven erweitert werden müssen. Mit dem Anstieg des technischen Erhebungsaufwandes verstärkt sich allerdings die unerwünschte Beeinflussung des zu erforschenden Feldes und der darin agierenden Personen durch so genannte Invasivitätseffekte (vgl. Petko et al. 2003: 270, Bortz/Döring 2006: 268), so dass die Datengrundlage zwar technisch an Qualität gewinnen mag, zugleich aber auf Grund möglicher Verzerrungen die Validität leiden kann. Abschließend kann die Aufbereitung der Daten je nach Aufnahmequalität, verfügbarer Hard- und Software und erwünschten Zieleigenschaften des Datenbestandes mit einem immensen Aufwand verbunden sein. Der Zugang videogestützter Beobachtungen von (Schul-) Unterricht ist somit zwar aus planungs- und organisationstechnischer Hinsicht voraussetzungsreich, gleichzeitig aber hinsichtlich methodischer als auch theoretischer Möglichkeiten äußerst vielversprechend. In der zeitgenössischen empirischen Unterrichtsforschung gehören Verfahren der systematischen Beobachtung unter Nutzung audiovisueller Aufnahmen mittlerweile zum gängigen Methodenrepertoire (vgl. Böhm-Kasper/Weishaupt 2004, Klieme 2006). Besonders durch Forschungsprojekte wie TIMSS[1], DESI (Helmke et al. 2008), der international-schweizerischen Videostudie (Reusser et al. 2010),

1 http://timssvideo.com/timss-video-study

der IPN-Videostudie[2] oder aktuell auch PERLE[3] und IGEL[4] hat die Verbreitung dieses Ansatzes im deutschsprachigen Raum in den letzten Jahren stark zugenommen, sodass auch methodologische Überlegungen, wie zum Beispiel zur Rolle der Videoforschung in gemischten Forschungsdesigns, verstärkt Eingang in den hiesigen wissenschaftlichen Diskurs finden. Möchte man den Erfolg oder die Effektivität der Lernumgebung oder des Lehrerhandelns bestimmen, dann reicht es nicht aus, diskursanalytisch den Text der Hauptakteure zu betrachten oder lediglich Testdaten und Unterrichtsmethoden in Beziehung zu setzen. Neuere Studien nutzen eine Vielzahl von Variablen der Instruktionsqualität (z.B. *Time on Task*), der Schülereigenschaften (z.B. das sprachliche Vorwissen) und des Settings (z.B. das Fachwissen der Lehrkraft) gleichzeitig, die entweder über Beobachtungsverfahren, Tests oder Befragungen erhoben worden sind. Diesbezüglich zeigt Abb. 1 beispielhaft ein Modell, in dem sich das Potential der Kombination von Erhebungsverfahren unter Einbezug von Videoanalysen andeutet: Dargestellt wird das Bedingungsgefüge zur Förderung der Lesefähigkeit auf unterschiedlichen Ebenen nach Connor et al. (2009: 87), welches besonders im Bereich der *multiple dimensions of instruction* einen Zugang über Verfahren videogestützter Beobachtung anbietet.

Abb. 1: Modell der Unterrichtsbedingungen zur Entwicklung der Lesefähigkeit (Connor et al. 2009: 87)



2 http://www.ipn.uni-kiel.de/projekte/video/
3 http://www.perle-projekt.de/
4 http://www.idea-frankfurt.eu/wissen/projekte/projekt-igel

Im Folgenden werden Ansätze videobasierter Beobachtungsstudien von Unterricht unter Berücksichtigung sowohl grundlegender methodischer Aspekte als auch verschiedener Möglichkeiten zur Triangulation anhand von Beispielen dargestellt.

2. Beobachten als sozialwissenschaftliches Verfahren der Datengewinnung

Triangulation kann in der erziehungswissenschaftlichen Forschung Verschiedenes bedeuten (vgl. Flick 2011). In diesem Artikel wird darunter in Anlehnung an die bis heute anerkannte Systematisierung von Denzin (1978) hauptsächlich eine Triangulation zwischen verschiedenen Datenquellen und Methoden ihrer Auswertung (*between-method*) verstanden. Diese Verfahrensweise dient unterschiedlichen Zwecken, wie im weiteren Verlauf noch an Beispielen zu zeigen ist.

In der Regel werden fünf quantitative Methoden der Datenerhebung unterschieden (vgl. Bortz/Döring 2006):

- Zählen (z.B. Index, Quantitative Inhaltsanalyse)
- Testen (z.B. Leistungstest, Persönlichkeitstest, usw.)
- Befragen (z.B. mündliches Interview, schriftlicher Fragebogen usw.)
- Physiologisches Messen (z.B. Pulsfrequenz etc.)
- Beobachten (z.B. Zeitstichprobe des Verhaltens).

In diesem Beitrag beschäftigen wir uns mit dem Beobachten im Rahmen von Videoprojekten. Mit diesem Verfahren werden Beobachtungsdaten produziert, die zu einem späteren Zeitpunkt z.B. mit Befragungs- oder Testdaten verknüpft werden können.

Grundsätzlich verstehen wir unter Verhaltensbeobachtung die auf das Verhalten eines oder mehrerer Organismen (Beobachtungsobjekte) gerichtete Wahrnehmung einer oder mehrerer Personen (Beobachter) mit dem Ziel, Wesentliches oder Symptomatisches über das Verhalten des bzw. der beobachteten Organismen oder über ihre Interaktion zu erfahren. Beabsichtigt ist die Herstellung eines möglichst intersubjektiv nachvollziehbaren Verhaltensprotokolls. In Videoanalyseprogrammen sind das in der Regel Datenreihen, die sich entlang einer Zeitachse anordnen lassen und mit Audio- oder Videosequenzen verbunden sind.

Worin sich unsere Alltagsbeobachtung von wissenschaftlichen Beobachtungen unterscheidet, wird an vielen Stellen (Gerrig/Zimbardo 2008) ausführlich dargelegt. Während Alltagsbeobachtung nach situativen und individuellen Präferenzen mehr oder weniger beliebig vonstatten geht, setzt die systematische Beobachtung einen genauen Beobachtungsplan voraus, der vorschreibt

- was (und bei mehreren Beobachtern auch von wem) zu beobachten ist,
- was für die Beobachtung unwesentlich ist,

- wann und wo die Beobachtung stattfindet und wie das Beobachtete zu protokollieren ist (z.B. live oder durch Video),
- ob bzw. in welcher Weise das Beobachtete gedeutet werden darf (angestrebter Grad der Trennung von Daten und Bewertungen).

Videografie ist eine Variante der instrumentellen Beobachtung, die zwischen den Daten, die ein Aufnahmeteam (Video- und Audiodaten, Transkripte, Materialien: Arbeitsblätter, Tafelbilder etc.) – auch hier schon selektiv (s.o.) – erfasst, und den Auswertungen durch Beobachter im Nachgang der Erhebung unterscheidet. In der Regel handelt es sich aus Gründen der ökologischen Validität[5] um Feldbeobachtungen in typischen Settings (z.B. Unterricht im Klassenzimmer), möglich wären aber auch experimentelle Designs im Labor.

Grundsätzlich unterscheiden wir bei der Auswertung zwischen Verfahren des *Beurteilens* oder des *Kodierens*.

2.1 Beurteilungsverfahren oder Ratings

Für den ersten Zugang werden sogenannte Ratingskalen genutzt, welche die nachträgliche Einschätzung eines Ereignisses oder Zeitraums anhand von Items mit Antwortvorgaben erlauben. Im Grunde unterscheidet sich das Procedere nur wenig von ähnlichen Verfahren der teilnehmenden Beobachtung. Vorteilhaft ist hierbei jedoch, dass mehrere Beobachter unabhängig von einander das gleiche Video einschätzen können, so dass im Unterschied zur Livebeobachtung die Interrater-Reliabilität, das heißt die Übereinstimmung zwischen Urteilen verschiedener Personen (vgl. Wirtz/Caspar 2002), relativ leicht bestimmt werden kann. Darüber hinaus eignet sich diese Vorgehensweise aufgrund ihrer Praktikabilität zur Analyse von großen Stichproben, bringt allerdings auch gewisse Schwächen mit sich, die an einem Beispiel verdeutlicht werden sollen.

Beispiel 1: Die Beurteilung von Unterrichtsqualität
Das folgende Item (Tab. 1) entstammt einem Ratingbogen zur Beurteilung von Aspekten der Unterrichtsqualität, welcher u.a. im Rahmen der so genannten „Münchener Hauptschulstudie“ (Helmke/Schrader 1993, 1997) und später in modifizierter Form in der DESI-Studie verwendet wurde (Helmke et al. 2008).

Ratings sind zwar relativ einfach anwendbar und bei hinreichender Schulung der Rater auch mit guter Übereinstimmung zu erstellen, haben aber den

---

5 Die so genannte ökologische Validität steht für den Grad der Generalisierbarkeit der Untersuchungsergebnisse aufgrund der „Natürlichkeit der Untersuchungssituation“ (Bortz/Döring 2006: 727).

Nachteil, dass komplexe Beurteilungsvorgänge der Beobachter in einem einzigen Wert verdichtet sind. Es ist für Dritte anhand der protokollierten Daten nicht rekonstruierbar, weshalb die Äußerungen einer Lehrkraft in einer bestimmten Unterrichtssequenz (siehe Beispiel) mit „geringe Verständlichkeit" und nicht mit „mäßige Verständlichkeit" bewertet wurden.

Tab. 1: Ausschnitt aus einem Ratingbogen zur Unterrichtsqualität (Helmke/Schrader 1993, 1997)

7. Verständlichkeit von Lehreräußerungen Verständlichkeit bezieht sich darauf, inwieweit die Darstellung des Lehrstoffs durch den Lehrer und seine sonstigen fachliche Äußerungen von den Schülern verstanden wurden. Wie verständlich waren die Erläuterungen des Lehrers?	
sehr hohe Verständlichkeit	Die Erklärungen des Lehrers waren leicht zu verstehen und die Fragen der Schüler wurden angemessen beantwortet. Der Lehrer schien die unterschiedlichen kognitiven Fähigkeiten der Schüler zu kennen. Er bemühte sich erfolgreich, Stoff zu erklären, mit dem die Schüler Schwierigkeiten hatten.
hohe Verständlichkeit	Zwischen 1. und 3. Kategorie.
mäßige Verständlichkeit	Der Lehrer schien von den meisten Schülern verstanden zu werden, wenn auch nicht die ganze Zeit über. Manchmal wirkte der Lehrer verwirrend und undeutlich.
geringe Verständlichkeit	Zwischen 3. und 5. Kategorie
sehr geringe Verständlichkeit	Die Schüler schienen durch die Darstellung des Lehrers sehr verwirrt zu werden. Der Lehrer konnte die Fragen der Schüler nicht erfolgreich beantworten oder er beantwortete sie unklar, indem er Begriffe oder Fachausdrücke verwendete, mit denen die Schüler offenbar nicht vertraut waren, oder indem er sich übermäßig kompliziert und undeutlich ausdrückte.

2.2 Kodierverfahren

Kodierverfahren setzen im Unterschied zu Ratings auf die möglichst genaue Protokollierung der vorab theoretisch definierten Ereignisse, Abläufe oder Interaktionen. Mit ihnen wird die Entwicklung von reproduzierbaren Beobachtungsprozeduren angestrebt, die den Beurteilungsspielraum der Beobachter stark eingrenzen. Grundsätzlich können (a) Zeichensysteme und (b) Kategoriensysteme unterschieden werden. Beide Systeme können entweder über Ereignis- oder Zeitstichproben Daten erfassen.

(a) Zeichensysteme für Ereignisstichproben oder Zeitstichproben
Kodieren kann durch einfache Zeichensysteme erfolgen. Darunter versteht man Listen von vorgegebenen Ereignissen, die bei Auftreten mit oder ohne Dauer notiert (z.B. Melden, Störungen) und auf der Videozeitachse festgehalten werden.

Zeichensysteme ermöglichen die genaue Beobachtung einer großen Anzahl verschiedener, singulärer und exakt definierter Verhaltensweisen. Sie können dann verwendet werden, wenn die zu beobachtenden Verhaltensweisen als individuelle Ereignisse (*Ereignisstichprobe*) sequenziell auftreten.

Beispiel 2: Die Erfassung von Sprachanteilen
Das Fundament von Unterricht basiert auf Kommunikation und Interaktion zwischen den Beteiligten, sodass dessen Erforschung entsprechende Aspekte berücksichtigen muss (vgl. Lüders 2003). Nicht nur aus Sicht der Fremdsprachendidaktik ist dabei die verbale Kommunikation von Interesse. Zugänglich wird dieser Bereich durch Audioaufnahmen, die dann entweder in schriftliche Transkriptionen überführt und dadurch für weitere, textanalytischen Methoden verfügbar gemacht werden (vgl. Wernet 2009), oder durch die direkte Verbindung der Aufnahmen mit Videoaufzeichnungen derselben Situation. Mit dem letztgenannten Verfahren sind verschiedene Vorteile verbunden, auch im Hinblick auf die Triangulation. Die Lehrer- und Schülersprache im Unterricht kann als sequenzielle Abfolge von einzelnen Sprechakten interpretiert werden, die durch eine Ereignisstichprobe anhand der audiovisuellen Aufnahmen erfassbar sind. Dies lässt sich unter Verwendung gängiger Analyse- und Auswertungsprogramme komfortabel durchführen[6]. Das Ergebnis ist ein präzises sequenzielles Protokoll der Sprachanteile inklusive ihrer Anfangs- und Endzeitpunkte, technisch verknüpft mit dem Videodokument (vgl. Tab. 2).

Tab. 2: Datenstruktur einer Ereignisstichprobe zu Sprachäußerungen im Unterricht

Anfang	Ende	Ereignis
00:00:10:04	00:00:13:09	Sprache Lehrperson
00:00:13:09	00:00:15:01	Sprache Lehrperson
00:00:15:05	00:00:16:17	Sprache Schueler/in1
00:00:16:17	00:00:18:06	Sprache Schueler/in2
00:00:18:23	00:00:21:08	Sprache Schueler/in3
00:00:25:13	00:00:27:13	Sprache Schueler/in4
00:00:27:13	00:00:28:16	Sprache Lehrperson
00:00:29:01	00:00:29:20	Sprache Schueler/in2
00:00:30:04	00:00:36:12	Sprache Schueler/in5
00:00:36:22	00:00:40:14	Sprache Schueler/in6

Anm.: Zeitangaben im Format: Stunden:Minuten:Sekunden:Bild-Frames

6 Z.B. mit dem Zusatzmodul Soundalyzer der Software Interact (Mangold International) vollständig automatisiert.

Zeichensysteme sind sehr einfach zu konstruieren, da man lediglich eine Liste der erwarteten Ereignisse und eine entsprechende Verfahrensvorschrift benötigt, welche die Ereignisse identifizierbar macht. Diese Ereignisse bzw. Verhaltensweisen (*Ereignisstichprobe*) können nun in weiteren Schritten beliebig oft nach unterschiedlichen Kriterien mit Bedeutung belegt bzw. klassifiziert werden – sei es hinsichtlich formaler Sprachmerkmale (Deutsch/Englisch/Französisch/etc.) oder des pädagogischen Handelns (Störungsintervention, Adressat in der Gruppe, Aktivitätsaufforderung etc.). Zeichensysteme haben jedoch den Nachteil, dass die erfassten Daten nur eingeschränkt interpretiert werden können. So lassen z.B. die Sprachereignisse nicht unmittelbar eine Berechnung des Redeanteils der Personengruppen am Unterricht zu, da die Ereignisse auch parallel stattfinden können (manchmal reden mehrere Personen im Klassenzimmer gleichzeitig), d.h. sie sind nicht notwendig exklusiv erfasst. Ein bestimmter Teil der Zeit wird zudem nicht erfasst, weil keine Ereignisse der definierten Liste stattfinden, d.h. die Kodierung ist nicht exhaustiv. Will man dagegen den Ereignisraum in einer Dimension vollständig darstellen, um auch relative Anteile zu bestimmen und weitere Vergleiche mit den Daten durchführen zu können, dann sind Kategoriensysteme die bessere Wahl.

(b) Kategoriensysteme

Kodiert wird dann durch kategoriale Zuordnung von Ereignissen oder Zeiträumen zu vorab theoretisch bestimmten Konstrukten (z.B. Unterrichtsphasen, Unterrichtsstile, Reaktionsweisen). Zu diesem Zweck werden Kategoriensysteme unter Beachtung dreier Prinzipien erzeugt (Bortz/Döring 2006: 140): (1) Genauigkeit, (2) Exklusivität und (3) Exhaustivität. Gemeint sind damit theoretische Begriffssysteme, die den Anspruch auf eine vollständige (3) Kodierung aller erwartbaren Verhaltensweisen in einer bestimmten Situation haben und gleichzeitig eine eindeutige (1) und überlappungsfreie (2) Kodierung der Objekte vornehmen. Ein einfaches Kodiersystem für rote, grüne und blaue Bälle wäre demnach nur solange eindeutig, wie es im Vorrat der vorhandenen Bälle keine rot-grün-gestreiften gäbe. Kategoriensysteme sind deshalb selten universell einsetzbar, sondern sind meistens an bestimmte Spezifika des Verhaltens in der jeweiligen Situation und Institution gebunden.

Beispiel 3: Phasen des Unterrichts

Oftmals werden Merkmale des Unterrichtsgeschehens in den Blick genommen, die kaum dem Verhalten einzelner Personen im Klassenraum zuzuordnen und eher übergeordneter Natur sind. Besonders die Frage nach der Lage und Dauer von bestimmten Unterrichtsphasen wird immer wieder in

verschiedenen Zusammenhängen gestellt, etwa, um unterschiedliche Muster des Unterrichtsverlaufs zu identifizieren oder spezifische Phasen für genauere Betrachtungen herauszugreifen (vgl. Seidel 2003a). Soll zum Beispiel die Segmentierung des Unterrichts aufgezeigt werden, ist es bedeutsam, die genauen Anfangs- und Endpunkte der unterschiedlichen Phasen zu kennen. Dies ist durch eine Ereignisstichprobe mit Dauer erreichbar (vgl. Seidel et al. 2003). Die Schwierigkeit bei diesem Beobachtungsplan besteht darin, die fokussierten Ereignisklassen vorab so präzise zu beschreiben, dass sowohl Anfangs- als auch Endpunkte für den Beobachter klar und unmissverständlich erkennbar sind, um die Ereignisse definieren zu können. Werden die Ereignisse innerhalb dieser erfassten Zeiträume kategorial kodiert, sind zudem die beschriebenen Prinzipien zu erfüllen. In unserer Studie wenden wir ein Beobachtungssystem zur Abbildung von Phasen mit verschiedenen Sozialformen an, welches dies veranschaulicht. Es werden folgende Kategorien unterschieden (Tab. 3):

Tab. 3: Kategoriensystem zur Bestimmung von Sozialformen im Unterricht

Kategorie	Kurzbeschreibung
Öffentlicher Unterricht (ÖU)	primärer Aufmerksamkeitsfokus bei Lehrperson oder legitimem Sprecher
Einzelarbeit (EA)	Schüler beschäftigen sich alleine mit Arbeitsauftrag; keine öffentliche Kommunikation
Partnerarbeit (PA)	Schüler beschäftigen sich zu zweit mit Arbeitsauftrag; keine öffentliche Kommunikation
Gruppenarbeit (GA)	Schüler beschäftigen sich in Gruppen (G >1) mit Arbeitsauftrag; keine öffentliche Kommunikation
Uneinheitliche Sozialform	Parallele Nutzung von PA, EA oder GA zum gleichen Zeitpunkt

Tab. 4: Datenstruktur einer Kodierung von Sozialform-Phasen aus einer beispielhaften Unterrichtsstunde (Format der Zeitangaben: hh:mm:ss)

Anfang	Ende	Sozialform
00:00:10	00:06:48	ÖU
00:06:48	00:13:11	GA
00:13:11	00:26:26	ÖU
00:26:26	00:36:27	EA

Im Falle der Kategorie des „öffentlichen Unterrichts“ sind Anfangs- und Endzeitpunkte aufgrund von verbalen oder non-verbalen Signalen der Lehrperson in der Regel gut definierbar, wohingegen besonders jene Phasen, die eine Veränderung des Klassen-Settings implizieren (Platzwechsel etc.), schwer zu verorten sind, da sie meist mit einer Art organisatorischer Übergangsphase verbunden sind und klare Signale der Lehrkraft mitunter ausbleiben. Weitere mögliche Konflikte des Kodierens können auch in der Abgrenzung zwischen den Kategorien (z.B. Partner- vs. Gruppenarbeit) entstehen (Prinzip der Exklusivität, s.o.). Daher sind genaue Kriterien erforderlich, die in schriftlichen Kodieranweisungen oder einem Beobachterhandbuch fixiert sein sollten. Außerdem sollten mehrere Kodierpersonen eingesetzt werden, um Beobachterübereinstimmungen sowohl hinsichtlich der Vergabe der Kategorien, als auch der Setzung der Zeitmarken für Anfangs- und Endpunkte[7] prüfen zu können. Sind diese Dinge sicher gestellt und die Daten erhoben, liegt schließlich eine wie in Tab. 4 abgebildete Datenstruktur vor.
Diese Angaben können nun als Ausgangspunkt für vertiefte Analysen dienen, um etwa bestimmte Segmente genauer in den Blick zu nehmen, was im Folgenden anhand eines Beispiels verdeutlicht wird. Des Weiteren können sie in Form von aggregierten deskriptiven Statistiken über mehrere Unterrichtsstunden oder Schülergruppen hinweg direkt in Auswertungen überführt werden.

Über diese Herangehensweise hinaus kann man Beobachtungsdaten auch als zusammenfassende Bewertung einer Menge von Ereignissen über einen längeren Zeitraum (*Zeitstichprobe*) gewinnen. Das ist immer dann sinnvoll, wenn das Geschehen von überlappenden, komplexen Ereignissen oder Verhaltensweisen geprägt ist (z.B. Handlungen von 20 Schülern), die zu einem späteren Zeitpunkt mit anderen Daten des gleichen Zeitpunkts (z.B. den Aktionen des Lehrers) in Beziehung gesetzt werden sollen. Die gesamte Beobachtungszeit wird dazu in Intervalle eingeteilt und für jedes Intervall wird eine Kodierung pro Beobachtungseinheit durchgeführt. Intervallbasierte Zeitstichproben strukturieren die Ereignisse also nach einem festgelegten Zeitraster, das über den Beobachtungszeitraum gelegt wird. Damit werden Ereignisse geschaffen, die auf künstlichem Wege zeitlich parallelisiert sind, was die Erfassung und Auswertung komplexer Merkmale vereinfacht. Das tatsächliche Auftreten mit natürlichem Anfangs- und Endpunkt wird damit allerdings nicht abgebildet. Hier soll als Beispiel für die Beobachtung von Zeitstichproben die Variable *Time on Task* erläutert werden.

---

7 Zur Berechnung entsprechender Kennwerte sei an dieser Stelle auf das Werk von Wirtz/Caspar (2002) verwiesen.

Beispiel 4: Die Beobachtung von Time on Task
In Anlehnung an Bloom (1974) wird *Time on Task* als ein Maß für Engagement betrachtet, das Einblicke in das Tätigkeitsprofil von Schülern über den Stundenverlauf ermöglicht. Dahinter steht eine einfache Überlegung: Nur dann, wenn Schüler ein Lernangebot nutzen, also *on-task* sind, wird man intentionales Lernen erwarten dürfen. Bei einer Beschäftigung mit unterrichtsfernen Dingen (Störungen, Privatgespräche etc.) sollte dagegen die Lernzeit ungenutzt bleiben.

In unserer Forschung verwenden wir dazu ein selbst entwickeltes System zur Erfassung des Beteiligungsverhaltens der Schüler auf Individualebene mithilfe einer *Zeitstichprobe*. Eine *Zeitstichprobe* bietet den Vorteil, dass eine für alle beobachteten Personen gemeinsame Metrik geschaffen wird, die später eine Reihe von Analysen und Verknüpfungen mit anderen Zeitreihen erlaubt. Die *Zeitstichprobe* basiert in unserem Fall auf Intervallen von einer Minute, so dass zu jedem Schüler für jede Minute der in den Blick genommenen Phasen des Unterrichts eine Beurteilung des Beteiligungsverhaltens vorliegt. Die Entscheidung über die Länge des Intervalls ist durchaus heikel und muss sorgfältig geprüft werden. Wählt man ein sehr kurzes Intervall, dann steigt die Anzahl der notwendigen Beobachtungsdaten stark an, so dass der Erhebungsaufwand an die Grenze der Praktikabilität stößt. Wählt man ein zu langes Intervall, dann können Verhaltensweisen in einem Intervall so stark wechseln, dass eine einheitliche Bewertung des Intervalls unmöglich wird. Die Beurteilung geschieht in unserem System auf der Basis von fünf Indikatoren, die in der Summe das Gesamturteil *on-* oder *off-task* bestimmen (Tab. 5). Durch die Einbeziehung der Indikatoren soll unerwünschten Urteilsverzerrungen (z.B. durch „Abfärben“ eines Globaleindrucks auf Urteile zu Einzelmerkmalen) vorgebeugt werden. Das Entscheidungsschema ist auf allen Stufen dichotom, d.h. werden die Einschätzungen der Hilfsindikatoren mehrheitlich im Sinne von *on-* oder *off*-task bewertet, dann wird das Verhalten innerhalb des betreffenden Zeitintervalls entsprechend eingestuft.

Tab. 5: Beurteilungssystem zum Beteiligungsverhalten von Schülern im Unterricht

Verhaltensindikatoren	Blickrichtung	–	+
	Aktive Beteiligung	–	+
	Andere Tätigkeit	–	+
	Motorische Unruhe	–	+
	Kommunikation	–	+
Gesamturteil		off-task	on-task

Unser Vorgehen verlangt zwar eine Vielzahl an Kodierentscheidungen[8], erzeugt aber Daten, die nachträglich durch weitere Beobachter für jedes Intervall überprüft werden können (vgl. Tab. 6). Diese Datenstruktur bietet die Möglichkeit, Urteilsübereinstimmungen zwischen verschiedenen Beobachtern bis auf die Ebene der einzelnen Verhaltensindikatoren zu bestimmen. Auf diese Weise können Schwächen des Systems (z.B. wiederholte Differenzen in einem bestimmten Verhaltensindikator über verschiedene Beobachterpärchen und Unterrichtsstunden hinweg), aber auch Missverständnisse in der Beobachterschulung (erkennbar über wiederholte Differenzen bei einem Beobachter) identifiziert werden (vgl. Wirtz/Caspar 2002).

Tab. 6: Time on Task-Werte von fünf Schülern einer ausgewählten Unterrichtsphase (aus Platzgründen ohne Werte der fünf Hilfsindikatoren)

Int.-Anfang	Int.-Ende	S1	S2	S3	S4	S5
00:16:58	00:17:58	1	1	1	1	1
00:17:58	00:18:58	1	1	1	1	1
00:18:58	00:19:58	1	1	1	1	1
00:19:58	00:20:58	1	1	1	1	1
00:20:58	00:21:58	1	1	1	1	0
00:21:58	00:22:58	1	1	1	1	1
00:22:58	00:23:58	0	1	1	0	0
00:23:58	00:24:58	1	1	1	1	1
00:24:58	00:25:58	1	1	1	1	1
00:25:58	00:26:58	1	1	1	1	1
00:26:58	00:27:58	1	1	1	1	1
00:27:58	00:28:58	0	1	1	0	1
00:28:58	00:29:58	1	1	1	0	0
00:29:58	00:30:58	1	1	1	1	0
00:30:58	00:31:58	1	1	1	1	0

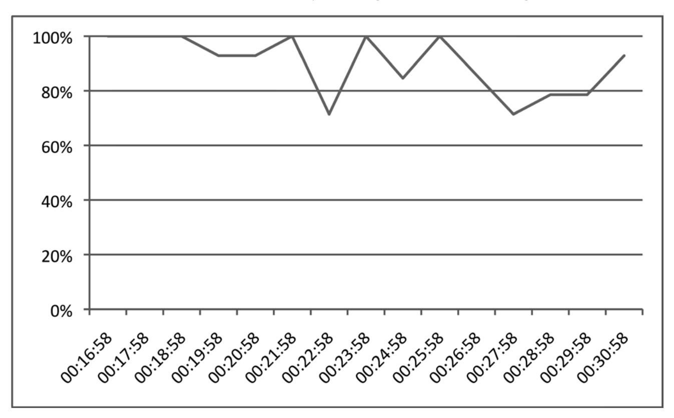
Anm.: 0 = off; 1 = on; Zeitangaben im Format: hh:mm:ss

8 Pro Minuten-Intervall und Schüler sind sechs Urteile zu fällen, d.h. beispielsweise bei 20 Schülern und einer 30-minütigen Phase 3600 Kodier-Entscheidungen.

3. Beispiele für die Auswertung von Kodierungen mit dem Ziel der Datentriangulation

Mit der dargestellten Datenbasis aus der oben erläuterten Kodierung der *Time on Task* lässt sich durch Aggregieren der Individualdaten auf Klassenniveau recht einfach ein Aktivitätsprofil über den Zeitverlauf erzeugen. Wir nennen das die „Reichweite der Lehrkraft" im jeweiligen Intervall. Unser System ist in der vorgestellten Form allerdings nur in Phasen des öffentlichen Unterrichts anwendbar, denn in anderen Phasen wären die Verhaltensindikatoren nicht plausibel. In einer Gruppenarbeitsphase ist z.B. der Blickkontakt mit dem Hauptakteur in jeder Gruppe anders definiert. Deshalb filtern wir mithilfe der Kategorien der Sozialform nur solche Daten, die in Phasen des „Öffentlichen Unterrichts (ÖU)" stattfinden[9]. Die Beobachtungsintervalle kann man grafisch vereinfacht als Zeitlinie (vgl. Abb. 2) visualisieren.

Abb. 2: Time on Task-Verlauf auf Klassenebene in einer ausgewählten Unterrichtsphase (Anteil der Schüler mit On-Task-Status im jeweiligen Beobachtungsintervall)



Verlaufsabbildungen wie diese lassen sich ebenso für Untergruppen innerhalb der Klasse erzeugen (z.B. nach Geschlecht, fachlichem Leistungsstand etc.). Aus theoretischer Sicht ist dies besonders für Fragen der Binnendifferenzierung oder Adaptivität bedeutsam, da sich dadurch Differenzen im Beteiligungsverhalten zwischen den verschiedenen Personengruppen zu bestimmten Zeitpunkten und deren Bezug zu Unterschieden in der Unterrichtsführung der Lehrkraft

9 Der durchschnittliche Anteil der Sozialform „ÖU" an der Unterrichtszeit beträgt in unserer Stichprobe 60 %

untersuchen lassen. Ebenso kann durch Aggregieren der Intervallwerte das personenspezifische On-Task-Niveau über die gesamte Stunde abgebildet werden (Tab. 7).

Tab. 7: Personenspezifisches On-Task-Niveau über alle beurteilten Intervalle in einer Unterrichtsstunde (ausgewählte Schüler)

Schüler-Nr.	S1	S2	S3	S4	S5	S6	S7	S8	S9
Anteil On-Task	76%	79%	83%	88%	68%	44%	58%	60%	100%

Der personenspezifische Stunden-Globalwert lässt sich wiederum auf Klassenebene zusammenfassen und kann dazu dienen, eventuelle Unterschiede zwischen Stunden oder Lerngruppen aufzuzeigen (Tab. 8).

Tab. 8: Durchschnittlicher Anteil an On-Task-Verhalten über alle Schüler in einer Klasse (in Prozent), über die Messzeitpunkte (Unterrichtsstunden)

	Klasse A	**Klasse B**
1. Std.	95 %	88 %
2. Std.	95 %	90 %
3. Std.	92 %	80 %
4. Std.	90 %	69 %
Durchschnitt	93 %	83 %

Auf dieser Ebene können, bei ausreichend großer Stichprobe, Unterschiede zwischen Lehrkräften oder Veränderungen in der Unterrichtsführung (z.B. im Rahmen von Interventionsstudien) in den Blick genommen werden. Zusätzlich eröffnen sich Möglichkeiten zur Daten-Triangulation, sprich der Kombination unterschiedlicher Datenquellen zum Zwecke höherer theoretischer Aussagekraft (vgl. Denzin 1978, Flick 2011: 13), indem die Angaben in Beziehung gesetzt werden zu Daten, die nicht durch Beobachtung oder Auswertung der Videoaufnahmen, sondern losgelöst von der Unterrichtssituation gewonnen werden, wie schriftliche Befragungen, Tests oder Interviews. Tab. 9 zeigt dazu beispielhaft Zusammenhänge der beobachteten Time on Task in Form der durchschnittlichen Rate über alle beobachteten Unterrichtsstunden auf Individualebene mit Testwerten aus einer schriftlichen Leistungsmessung und Hintergrunddaten der Personen, gewonnen in einer schriftlichen Befragung.

Tab. 9: Beziehungen der Variable Time on Task zu Befragungs-/Testdaten

Mittelwertvergleiche (Testvariable: Time on Task)		
	t-Test	Effekt (d)
Geschlecht (m/w)	t(130) = −3.29, p < .01	.58
Sprachhintergrund (deutsch/andere)	t(130) = 0.72, p > .05	.16
Zusammenhänge (mit Time on Task)		
	Zusammenhang*	
Testleistung	r = .12, p < .05	

*Produkt-Moment-Korrelation nach Pearson

Daran schließen sich weitere Möglichkeiten an, verschiedene Datenquellen zusammengefasst zu veranschaulichen. Am genannten Beispiel geschieht dies mit einer Art „multiplem Sitzplan", in dem die Beobachtungsdaten zum Verhalten und Befragungsdaten mit Personeneigenschaften, ergänzt durch beobachtungsbasierte Angaben über Wortmeldungen und Aufrufe, mit einem Sitzplan kombiniert werden (Abb. 3). Diese Darstellung kann zur Exploration der Datenlage und zur Hypothesengenerierung neuer Fragestellungen herangezogen werden – etwa hinsichtlich der Sitzordnung bzw. der räumlichen Anordnung.

Abb. 3: Sitzplan unter Einbezug von Test-, Personen- und Beobachtungsdaten

4. Zusammenfassung

In diesem Beitrag wurde systematisches Beobachten unter Verwendung von audiovisuellen Aufzeichnungen als eine Variante sozialwissenschaftlicher Datenerhebung vorgestellt. Der Schwerpunkt lag dabei auf besonderen Vorteilen videobasierter Beobachtungsverfahren zur Erforschung des Gegenstands schulischen Unterrichts. Das zentrale Argument für die Verwendung dieser Methode besteht darin, dass die lange Zeit gängigen Feldbeobachtungen im Klassenraum (in der Regel nicht-teilnehmend und offen) durch die hohe Komplexität des Geschehens an Grenzen gestoßen sind. In der Folge wurde oft auf Ratings zurückgegriffen, die allerdings nur schwach zwischen Beobachtung und Interpretation trennen. Videodaten bieten dagegen den Vorteil, dass sie die Anwendung komplexer Beobachtungssysteme erlauben. Im Weiteren wurden Vorzüge sowohl des wiederholten Zugriffs auf das Material als auch der Möglichkeit der Beteiligung mehrerer Forschungspersonen deutlich, zugleich aber auch damit verbundene methodische Erfordernisse, die in Form von Regeln oder Prinzipien zur Wahrung der Qualität im Forschungsprozess dienen (z.B. Auswahl der Beobachtungsstichprobe, Kategorienbildung, Beobachterübereinstimmung).[10]

Die Datenbeispiele haben gezeigt, dass durch diverse Arten der Erfassung und Protokollierung von Beobachtungsdaten (Ratings, Kodierung mit Ereignis- oder Zeitstichprobe, Kategoriensysteme) unterschiedliche Datenstrukturen geschaffen werden, die vielfältige Anknüpfungspunkte zur Verbindung mit alternativen Datenquellen oder auch der Auswertung auf verschiedenen Ebenen bieten. Diese Möglichkeiten erscheinen im Kontext des Begriffs der Triangulation als vielversprechend. Videobasierte Beobachtungsstudien entfalten ihr Potenzial allerdings für bestimmte Arten von Triangulation in besonderem Maße. Mit Blick auf die Unterscheidung von Denzin (1978) sind dabei zwei Formen hervorzuheben. Erstens die Methodentriangulation, speziell die so genannte *within-method triangulation* (ebd.: 301), d.h. eine Variation des Skalierungsverfahrens oder Beurteilungssystems innerhalb der Methode. Darunter können Maßnahmen zur Sicherung der Validität von Beurteilungsinstrumenten gefasst werden, wie etwa die Erprobung verschiedener Varianten von Ratingskalen in einem iterativen Prozess der Instrumentenkonstruktion (vgl. Seidel 2003b). Zweitens die Daten-Triangulation, also der Verbindung von Datenbeständen aus unterschiedlichen Erhebungsverfahren (Denzin 1978), der im Kontext der Videoforschung besonders relevanten Triangulationsform. Hierbei kommt der Vorteil zum Tragen,

---

10 Es sei darauf hingewiesen, dass der Schwerpunkt in diesem Text auf quantitativen Zugängen zum Datenmaterial lag. Darüber hinaus verweisen wir auf Herrle et al. (2010).

dass durch audiovisuelle Aufzeichnungen Aspekte zugänglich gemacht werden, die durch ihre Flüchtigkeit auf andere Weise nur schwer zu erfassen sind (beispielsweise konkrete Verhaltensweisen einzelner Personen). Die auf diese Weise erfassten Merkmale lassen sich wie dargestellt gut mit alternativ erhobenen Merkmalen kombinieren, was wiederum die Einordnung in komplexere und umfassendere Rahmenmodelle von Unterricht erlaubt (vgl. Abschnitt 1). Ferner bieten sich durch Bearbeitung des Materials z.B. in Forschungsverbünden auch gewisse Spielräume zur Theorie-Triangulation an, also der Annäherung an den Gegenstand aus unterschiedlichen theoretischen Perspektiven und mit variierenden Hypothesen (Denzin 1978), etwa durch kontrastierende oder ergänzende Beschreibungsansätze anhand von Interaktionstheorien, Theorien der pädagogischen Psychologie oder der Fremdsprachenforschung. Insgesamt sollte aber deutlich werden, dass videobasierte Beobachtungsstudien nicht nur Möglichkeiten zur Triangulation eröffnen, sondern sie in Gestalt der Verknüpfung und Kombination von Daten, der Variation von Instrumenten und der Beteiligung mehrerer Forscher/innen regelrecht erfordern, um Beobachtungsdaten in ihrem Aussagegehalt aufwerten zu können.

Die Umsetzung der hier dargestellten Verfahren würde ohne technische Hilfsmittel sehr schnell an Grenzen stoßen, sodass mittlerweile eine Reihe entsprechender Computerprogramme zur Analyse von Videodaten vorliegt. Durch differierende Schwerpunktsetzungen hinsichtlich Programmaufbau, Funktionen und interner Datenstruktur werden für die Nutzer jedoch jeweils unterschiedliche Analyse- und Auswertungsszenarien besonders begünstigt und andere teilweise eher erschwert. Sehr spezielle Kodiervorgänge wie beispielsweise das vorgestellte Kategoriensystem zum Beteiligungsverhalten von Schülern im Unterricht, bei dem eine zeitstichprobenbasierte Beobachtung zu mehreren unterschiedlichen Verhaltenskomponenten auf verschiedene Personen gleichzeitig angewendet wird, lassen sich nicht in allen Programmen gleichermaßen praktikabel umsetzen. Die erfolgreiche Implementierung von wissenschaftlichen Instrumenten und Arbeitsweisen in eine digitale Arbeitsumgebung setzt daher eine bewusste Auswahl der geeigneten Analyse-Software voraus.

Literatur

Adams, Raymond S. / Biddle, Bruce J. (1970). *Realities of teaching*. New York: Holt Rinehart Winston.

Aufschnaiter von, Stefan / Welzel, Manuela (2001). Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen: Eine Einführung. In: Aufschnaiter von, Stefan / Welzel, Manuela (Hrsg.). *Nutzung von Videodaten zur*

Untersuchung von Lehr-Lern-Prozessen. Aktuelle Methoden Empirischer Pädagogischer Forschung. Münster: Waxmann, 7-15.

Biddle, Bruce J. (1967). Methods and concepts in classroom research. *Review of Educational Research* 37/3, 337-357.

Bloom, Benjamin S. (1974). Time and Learning. *American Psychologist* 29/9, 682.

Bortz, Jürgen / Döring, Nicola (2006). *Forschungsmethoden und Evaluation*. 4. Aufl. Heidelberg: Springer.

Böhm-Kasper, Oliver / Weishaupt, Horst (2004). Quantitative Ansätze und Methoden in der Schulforschung. In: Helsper, Werner / Böhme, Jeanette (Hrsg.). *Handbuch Schulforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 93-125.

Connor, Carol M. / Morrison, Frederick J. / Fishman, Barry J. / Ponitz, Claire C. / Glasney, tephanie / Underwood, Phyllis S., Piasta, Shayne B. / Coyne Crowe, Elizabeth/ Schatschneider, Christopher (2009). The ISI Classroom Observation System: Examining the Literacy Instruction Provided to Individual Students. *Educational Researcher* 38/2, 85-99. doi:10.3102/0013189X09332373.

Denzin, Norman K. (1978). *The Research Act.*. 2.. Aufl. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Dinkelaker, Jörg / Herrle, Matthias (2009). *Erziehungswissenschaftliche Videografie*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick, Uwe (2011). *Triangulation - Eine Einführung*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gerrig, Richard J. / Zimbardo, Philip G. (2008). *Psychologie*. München: Pearson Deutschland.

Helmke, Andreas / Schrader, Friedrich-Wilhelm (1993). Was macht erfolgreichen Unterricht aus? Ergebnisse der Münchner Studie. *Praxis Schule* 5-10/1, 11-13.

Helmke, Andreas / Schrader, Friedrich-Wilhelm (1997). Unterrichtsbeurteilungen durch externe Beobachter. In: Weinert, Franz E. / Helmke, Andreas (Hrsg.). *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union, 510-518.

Helmke, Tuyet / Helmke, Andreas / Schrader, Friedrich-Wilhelm / Wagner, Wolfgang / Nold, Günter / Schröder, Konrad (2008). Die Videostudie des Englischunterrichts. In: Klieme, Eckhard (Hrsg.). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz, 345-363.

Herrle, Matthias / Kade, Jochen / Nolda, Sigrid (2010). Erziehungswissenschaftliche Videografie. In: Friebertshäuser, Barbara / Langer, Antje / Prengel, Annedore (Hrsg.). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. 3. Aufl. Weinheim: Juventa, 599-620.

Jacobs, Jennifer K. / Kawanaka, Takako / Stigler, James W. (1999). Integrating Qualitative and Quantitative Approaches to the Analysis of Video Data on Classroom Teaching *International Journal of Educational Research* 31/8, 717-724.

Klieme, Eckhard (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. *Zeitschrift für Pädagogik* 52/6, 765-773.

Kounin, Jacob (1976). *Techniken der Klassenführung*. Bern: Huber & Klett.

Lüders, Manfred (2003). *Unterricht als Sprachspiel. Eine systematische und empirische Studie zum Unterrichtsbegriff und zur Unterrichtssprache*. Bad Heilbrunn: Klinkhardt.

Pauli, Christine / Reusser, Kurt (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik* 52/6, 774-798.

Petko, Dominik / Waldis, Monika / Pauli, Christine / Reusser, Kurt (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *Zentralblatt für Didaktik der Mathematik* 35/6, 265-280.

Reusser, Kurt / Pauli, Christine / Waldis, Monika (Hrsg) (2010). *Unterrichtsgestaltung und Unterrichtsqualität*. Münster: Waxmann.

Seidel, Tina (2003a). Sichtstrukturen – Organisation unterrichtlicher Aktivitäten. In: Seidel, Tina / Prenzel, Manfred / Duit, Reinders / Lehrke, Manfred (Hrsg.). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse Im Physikunterricht"*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), 113-128.

Seidel, Tina (2003b). Videobasierte Kodierverfahren in der IPN Videostudie Physik – ein methodischer Überblick. In: Seidel, Tina / Prenzel, Manfred / Duit, Reinders / Lehrke, Manfred (Hrsg.). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse Im Physikunterricht"*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), 99-111.

Seidel, Tina / Prenzel, Manfred / Duit, Reinders / Lehrke, Manfred (Hrsg.) (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse Im Physikunterricht"*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).

Wernet, Andreas (2009). *Einführung in die Interpretationstechnik der Objektiven Hermeneutik*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Wild, Klaus Peter (2003). Videoanalysen als neue Impulsgeber für eine praxisnahe, prozessorientierte empirische Unterrichtsforschung. *Unterrichtswissenschaft* 31/2, 98-101.

Wirtz, Markus A. / Caspar, Franz (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe Verlag für Psychologie.

Videografie als Instrument zur Erforschung von Interaktionsprozessen im Elementarbereich

Heidi Seifert

Während im Bereich der empirischen Unterrichts- und Bildungsforschung ein allgemeiner Anstieg videobasierter Forschungsaktivitäten zu verzeichnen ist, wird die Videografie als Instrument zur Erforschung von Bildungsprozessen im Elementarbereich noch selten eingesetzt. Dies spiegelt sich auch in der überschaubaren Publikationslage zur methodischen Anlage von Videostudien im Elementarbereich wider. Der vorliegende Beitrag nimmt dies zum Anlass, um die Einsatzmöglichkeiten der Videografie im Forschungsfeld Kita darzustellen und unter verschiedenen Gesichtspunkten zu diskutieren. Am Beispiel einer Dissertationsstudie zu Erzieherin-Kind-Interaktionen in einer bilingualen Kita werden die einzelnen Phasen im videobasierten Forschungsprozess – von der Vorbereitung der Datenerhebung, die insbesondere die Beachtung datenschutzrechtlicher Aspekte umfasst, über die Durchführung der videobasierten Datenerhebung bis hin zur Aufbereitung und Analyse der Videodaten – explizit aufgegriffen und methodische Überlegungen eingehend diskutiert.

1. Einleitung

Das Verfahren der Videografie konnte sich in den letzten Jahren in der empirischen Unterrichts- und Bildungsforschung zunehmend etablieren. Auch in der Fremdsprachendidaktik findet die videobasierte Beobachtung von Lehr- und Lernprozessen verstärkt Beachtung. Die Zunahme videogestützter Forschungsaktivitäten ist in erster Linie auf technische Entwicklungen zurückzuführen, die gänzlich neue Perspektiven für die Videografie eröffneten. Während in den 70er und 80er Jahren die Durchführung von Videostudien mit einem erheblichen Kosten- und Planungsaufwand verbunden war, ermöglichen technische Neuerungen heutzutage eine vergleichsweise komfortable Aufzeichnung, Verarbeitung und Analyse audiovisueller Daten und führten so zu „einem Qualitätssprung der Videoaufzeichnung bei gleichzeitig sinkenden Kosten" (Petko et al. 2003: 265). Moderne Camcorder zeichnen in hochauflösendem Digitalformat auf und gewährleisten so ein hohes Maß an Bild- und Tonqualität. Zudem wurden Softwaretools entwickelt (z.B. *Videograph*, *Interact*, *Catmovie, Atlas.ti*), die eine deutlich verbesserte und komfortable Archivierung, Verarbeitung und Analyse selbst großer Datensätze ermöglichen.

Neben den technischen Entwicklungen, die zu einer zunehmenden Beliebtheit der Videografie geführt haben, zeigt ein Vergleich mit der verwandten Methode der teilnehmenden Beobachtung, dass der Einsatz der Videografie – in Abhängigkeit von dem jeweiligen Forschungsgegenstand und dem individuellen Erkenntnisinteresse – für einige Beobachtungsstudien einen forschungsmethodischen Mehrwert bedeuten kann. Während die teilnehmende Beobachtung teilweise von den theoretischen Vorannahmen des Forschers eingefärbt sein kann, ermöglicht die videogestützte Beobachtung eine „relativ theorieunabhängige und komplexitätserhaltende Dokumentation" (Schramm/Aguado 2010: 186), so dass Videodaten gegenüber Daten aus einer teilnehmenden Beobachtung eine „weniger subjekt- und theoriegebundene Qualität" (Petko et al. 2003: 265) aufweisen können. Ein weiterer Vorteil gegenüber der teilnehmenden Beobachtung besteht darin, dass die Beobachtung iterativ durchführbar ist, d.h. dass die Beobachtung beliebig oft wiederholt werden kann (vgl. u.a. Knoblauch/Schnettler 2007: 586). Dank der Reproduzierbarkeit der Beobachtung ist es fortan auch möglich, unterschiedliche Fragestellungen und Erkenntnisinteressen verschiedener Forscher/innen an einen einmal erhobenen Videodatensatz heranzutragen (vgl. Pauli/Reusser 2006: 791).

Die Videografie kann demnach als Methode beschrieben werden, die die begrenzte Wahrnehmungs- und Verarbeitungskapazität von Beobachter/innen im Forschungsfeld sowie die Flüchtigkeit des Geschehens überwinden kann. Die videobasierte Beobachtung gilt daher als Erhebungsinstrument, das der Komplexität von Unterrichtsprozessen bzw. Lehr- und Lernprozessen am ehesten Rechnung trägt (vgl. Schramm/Aguado 2010). Aufgrund ihrer dokumentierenden Funktion und Anschaulichkeit können Videodaten darüber hinaus für die Aus- und Weiterbildung von Lehrkräften nutzbar gemacht und als flexibles Medium zur Aneignung professionellen Wissens und professioneller Handlungsstrategien von Lehrkräften eingesetzt werden (zu verschiedenen Einsatzmöglichkeiten von Videos in der Lehrerbildung siehe Fischer/Schratz 2005). Videoaufzeichnungen, die im wissenschaftlichen Kontext erhoben wurden, können beispielsweise als Anschauungsmaterial zur Illustration von Forschungsergebnissen in der Lehrerbildung dienen und so eine „Brücke über die Kluft zwischen didaktischer Theorie und Praxis" (Petko et al. 2003: 278) schlagen sowie einen „Kristallisationspunkt der fachdidaktischen Verständigung" (ebd.: 278) bilden.

Parallel zum Anstieg videobasierter Forschungsaktivitäten im Bereich der empirischen Unterrichtsforschung fällt jedoch auf, dass Publikationen zum methodischen Design von Videostudien und zu Einsatzmöglichkeiten des Verfahrens der Videografie vorwiegend im mathematisch-naturwissenschaftlichen Bereich zu verorten sind (siehe z.B. Aufschnaiter/Welzel 2001, Hugener et al.

2006). In der (Fremd-)Sprachendidaktik scheinen Methodendiskussionen zur Durchführung von videogestützten Beobachtungsstudien bisher weniger etabliert zu sein. Hinsichtlich der defizitären Publikationslage zur methodischen Anlage von Videostudien stellen so auch Ricart Brede et al. (2010: 259) resümierend fest: „Insgesamt stellt die explizite Beschreibung der Anlage und Durchführung von Videostudien im Bereich des Sprachunterrichts, zumindest des Deutschunterrichts, ein Desiderat dar.“ Darüber hinaus spiegelt die Sichtung der Literatur zur Videografie wider, dass Videostudien bisher vorwiegend im Kontext schulischer Forschung zum Einsatz kommen. Für die videobasierte Erforschung von Lern- und Bildungsprozessen im Elementarbereich liegen zwar erste empirische Studien vor (z.B. König 2006, Ricart Brede 2011), jedoch scheint die Methode für diesen Bildungssektor bisher weniger Beachtung gefunden zu haben als in der empirischen Unterrichtsforschung.

Anliegen des vorliegenden Beitrags ist es daher, die Genese einer Videostudie im Elementarbereich offenzulegen und die methodischen Überlegungen, die bei der Entwicklung des Forschungsdesigns sowie der Planung und Durchführung der videobasierten Datenerhebung eingeflossen sind, zu diskutieren. Dies geschieht am Beispiel eines Dissertationsprojekts zum sprachlichen Interaktionsverhalten von pädagogischen Fachkräften in einer bilingualen Kita. Der Forschungsgegenstand und die Ziele der Studie werden im Folgenden skizziert.

2. Erkenntnisinteresse des Forschungsvorhabens

Die europäische Sprachenpolitik setzt sich angesichts der zunehmenden Globalisierung und Internationalisierung des Arbeitsmarktes für die Förderung der individuellen Mehrsprachigkeit, den Erhalt und die Pflege der Sprachenvielfalt in Europa, sowie ein lebenslanges Fremdsprachenlernen ein (vgl. Europäische Kommission 1996). Insbesondere die Umsetzung der so genannten 3-Sprachenformel (Erstsprache und zwei Fremdsprachen), die es jedem EU-Bürger ermöglichen soll, im Laufe seines Lebens Kompetenzen in mehreren Sprachen zu erwerben, ist ein erklärtes Ziel der europäisches Sprachenpolitik (vgl. ebd.). In den letzten Jahren fanden die sprachenpolitischen Empfehlungen durch die Einrichtung bilingualer Grundschulklassen und der Eröffnung bilingualer Kindergärten erste Umsetzungen. In Deutschland steigt vor allem die Nachfrage an bilingualen Kitas, so dass ihr Anteil allein seit 2007 um 25% angestiegen ist (vgl. FMKS 2010). Einrichtungen mit der Sprachenkombination Deutsch-Englisch nehmen mit 42% eine dominante Stellung unter den bilingualen Vorschulangeboten ein (vgl. ebd.). Viele dieser zweisprachigen Kindergärten arbeiten nach der Methode der Immersion, die aufgrund der Erkenntnisse der kanadischen

Immersionsforschung als eine der erfolgreichsten Methoden der Fremdsprachenfrühvermittlung gilt (vgl. u.a. Genesee 1987, Wode 1995, Zydatiß 2000). Im Rahmen dieser Methode ist die zu erlernende Sprache nicht der Lerngegenstand wie im traditionellen Fremdsprachenunterricht, sondern die Arbeitssprache, die sämtliche Aktivitäten und den gesamten Tagesablauf in der Einrichtung begleitet. Die Sprachverwendung erfolgt innerhalb der Methode der Immersion nach der funktionalen Sprachentrennung gemäß dem Prinzip *one person-one language* (vgl. Döpke 1992). So interagieren beispielsweise in einer deutsch-englischen Kita die deutschsprachigen Fachkräfte ausschließlich auf Deutsch und die englischsprachigen Fachkräfte ausschließlich auf Englisch mit den Kindern.

Findet kindlicher Spracherwerb in einer bilingualen Kita in einem immersiven Kontext und gemäß der personengebundenen Sprachverwendung statt, kommt der sprachlichen Interaktion zwischen den Kindern und den pädagogischen Fachkräften eine wesentliche Bedeutung für den bilingualen Spracherwerb zu. Die Bedeutung der sozialen Lernumwelt für die Aneignung sprachlicher Kompetenzen greifen sowohl interaktionistisch-kognitivistische als auch soziokulturelle Erklärungsansätze zum Spracherwerb auf wie die Interaktionshypothese nach Long (1996) oder die „Zone der nächsten Entwicklung“ nach Wygotski (1987). Vor dem Hintergrund dieser interaktionistisch-kognitivistisch und soziokulturell geprägten Annahmen zum Spracherwerb kommt pädagogischen Fachkräften in Kitas eine tragende Rolle als Interaktions- und Kommunikationspartner zu, welche die sprachlichen Entwicklungschancen der betreuten Kinder durch ihr sprachliches Interaktionsverhalten entscheidend beeinflussen können.

Trotz der steigenden Beliebtheit bilingualer Kitas sind die spezifischen Rahmenbedingungen, denen der kindliche Spracherwerb in jenen Einrichtungen unterliegt, nach wie vor weitestgehend unerforscht. Das Dissertationsprojekt knüpft an dieses Forschungsdesiderat an, indem es in der deutsch-englischen Kita der TU Darmstadt ausgehend von interaktionistischen und soziokulturellen Erklärungsansätzen zum Spracherwerb Interaktionsprozesse zwischen pädagogischen Fachkräften und Kindern untersucht. In der Einrichtung, in der derzeit 40 Kinder von Bediensteten und Studierenden der TU Darmstadt betreut werden, wird die beschriebene Methode der Immersion und die personengebundene Sprachverwendung als Spracherziehungskonzept eingesetzt. Die leitende Frage des Dissertationsvorhabens ist, wie die deutsch- und englischsprachigen pädagogischen Fachkräfte im Kontext von Immersion und dem Prinzip *one person-one language* folgend sprachliche Interaktionen mit den ein- bis dreijährigen Kindern gestalten und welche Charakteristika das sprachliche Verhalten in diesen Situationen aufweist. Diese übergeordnete Forschungsfrage wurde in die folgenden Teilfragen aufgeschlüsselt.

1. Welche Arten von Erzieherin-Kind-Kommunikationssituationen kommen während der Freispielzeit in der bilingualen Kita zustande?
2. Welche Charakteristika weist das sprachliche Verhalten der Erzieherinnen auf und welche Formen sprachlicher Unterstützung leisten sie in den verschiedenen Interaktionssituationen?

Mittels der ersten Teilfrage sollen Aktivitäten identifiziert werden, die rahmengebend und Voraussetzung für sprachliche Interaktionsprozesse zwischen pädagogischen Fachkräften und Kindern sind. Darauf aufbauend sollen in einem zweiten Schritt mikroanalytisch einzelne Sequenzen näher in den Blick genommen und das sprachliche Interaktionsverhalten der Erzieherinnen mit einem besonderen Fokus auf den Einsatz von Strategien des *Scaffoldings* und Aspekten der Responsivität analysiert werden.

Die Ergebnisse der Studie werden dezidiert beschreiben, wie sprachliche Interaktionsprozesse zwischen pädagogischen Fachkräften und Kindern in der bilingualen Kita gestaltet sind. Zugleich sollen sie einen empirisch abgesicherten Einblick in die sprachliche Bildungsarbeit bilingualer Kitas liefern, der bislang in der Form noch nicht vorliegt. Damit können die Ergebnisse einen Beitrag zur Erforschung von Bedingungsfaktoren des kindlichen Spracherwerbs in bilingualen Kitas leisten. Darüber hinaus können die wissenschaftlichen Erkenntnisse der Untersuchung auch eine Relevanz für die Praxis haben, indem sie als Ausgangsbasis für die Entwicklung von Weiterbildungs- und Qualifizierungsmaßnahmen für pädagogisches Personal in bilingualen Kitas dienen können. In diesem Zusammenhang können auch die Videodaten zur Veranschaulichung und Reflexion eingesetzt werden.

3. Der Einsatz von Videografie in der vorliegenden Studie

Ausgehend vom Forschungsgegenstand des Dissertationsprojekts soll in diesem Kapitel erörtert werden, warum die videobasierte Beobachtung eine geeignete Methode zur Bearbeitung der Forschungsfragen darstellt und wie sie in der vorliegenden Studie eingesetzt wird. Im Folgenden soll zunächst beschrieben werden, wie der Entscheidungsprozess für diese Methode in Verbindung mit der Generierung der Forschungsfragen gestaltet war. Darauf aufbauend sollen Punkte thematisiert werden, die bei der Vorbereitung einer videobasierten Datenerhebung im Elementarbereich zu beachten sind. Es werden zentrale Aspekte der Hauptdatenerhebung skizziert, bevor das Kapitel mit einem Ausblick auf das weitere geplante methodische Vorgehen, insbesondere bei der Datenauswertung, schließt.

3.1 Vorüberlegungen

Vor der eigentlichen Datenerhebung wurde das Forschungsfeld mehrere Monate mittels teilnehmender Beobachtung erkundet. Durch die regelmäßigen, d.h. an zwei Vormittagen pro Woche stattfindenden Beobachtungen konnte ein erster Feldzugang geschaffen und ein Vertrauensverhältnis zu den pädagogischen Fachkräften und der Kita-Leitung, aber auch zu den Kindern und deren Eltern aufgebaut werden. Die teilnehmende Beobachtung ermöglichte zudem wertvolle erste Einblicke in den Kita-Alltag mitsamt seiner Routinen, die Außenstehenden in dieser Form nicht zugänglich wären (vgl. Bortz/Döring 2006: 267).

Die Erkundung des Forschungsfeldes kann in diesem Forschungsvorhaben als Prozess verstanden werden, der sich in seinen Grundzügen an den „klassischen" Phasen einer teilnehmenden Beobachtung orientiert, die von Spradley (1980, zitiert nach Flick 2004: 207) beschrieben wurden. Diese Phasen sind gekennzeichnet durch eine zunehmende Konkretisierung und Konzentration auf einen speziellen Forschungsgegenstand (vgl. Flick 2004: 207). Die teilnehmende Beobachtung erfolgte dabei zunächst als „deskriptive" (ebd.) oder auch als „freie", „offene", „unstandardisierte" (Bortz/Döring 2006: 269) Beobachtung. Es sollten in einem ersten Schritt, einem qualitativ-explorativ orientierten Forschungsdesign Rechnung tragend, unvoreingenommen alle Aspekte in die Beobachtung einbezogen werden, um dadurch die Komplexität des Forschungsfeldes möglichst differenziert erfassen zu können. Die Eindrücke aus der teilnehmenden Beobachtung sowie umfassende Ereignisabläufe und interessant erscheinende Einzelheiten wurden nach jedem Kita-Besuch in einem Forschungstagebuch festgehalten. Parallel zur teilnehmenden Beobachtung und der Verschriftlichung der Beobachtungsergebnisse im Forschungstagebuch wurde die Forschungsliteratur gesichtet und mit den Eindrücken aus dem Kita-Alltag abgeglichen. Durch diesen zyklischen Prozess verengte und verdichtete sich der Beobachtungsfokus zunehmend zu einer „fokussierten Beobachtung" (Spradley 1980, zitiert nach Flick 2004: 207) der sprachlichen Interaktionsprozesse zwischen den pädagogischen Fachkräften und den Kindern während der Freispielzeit. Das endgültige Forschungsinteresse wurde schließlich in den oben genannten Forschungsfragen konkretisiert und in mehrere Teilfragen gegliedert. Die teilnehmende Beobachtung fand in der Folgezeit zunehmend „selektiver" statt (ebd.) und konzentrierte sich größtenteils auf Beobachtungssituationen, die für die formulierte Fragestellung relevant waren.

Im Zuge der Generierung der Forschungsfragen und der Verdichtung des Beobachtungsfokus auf die sprachlichen Interaktionsprozesse zwischen Erzieherinnen und Kindern erfolgten zudem wichtige Vorüberlegungen zum Einsatz der

Videografie als Datenerhebungsinstrument in der vorliegenden Studie. Zwar fungierte die teilnehmende Beobachtung in der Kita als wichtiger Schlüssel, um einen ersten Feldzugang zum Forschungsfeld und den Akteuren herzustellen und um direkt aus dem Feld die Forschungsfragen zu entwickeln, doch erwies sich die Methode als unzulänglich, um einen angemessenen Datenkorpus für die Dissertationsstudie zu erstellen. Aufgrund der Komplexität und Vielschichtigkeit des Forschungsgegenstandes – sprachliche Interaktionen enthalten sowohl verbale, paraverbale und nonverbale Elemente – stellte es sich äußerst schwierig dar, diese Fülle an Informationen zu protokollieren. Zudem konnte meistens erst im Anschluss an die Beobachtung und nicht direkt in der Situation protokolliert werden. Da Erinnerungen an konkrete Situationen jedoch nur zeitlich begrenzt abrufbar sind, waren die Protokolle oft lückenhaft und anfällig für Beobachtungsfehler und subjektive Verzerrungen. In Verbindung mit der Flüchtigkeit und mangelnden Konservierbarkeit der Daten aus der teilnehmenden Beobachtung wurde es zudem als zunehmend unbefriedigend empfunden, dass die Daten in ihrer Authentizität und Informationsfülle nicht unmittelbar mit anderen Forscher/innen zu Zwecken der kommunikativen Validierung geteilt werden konnten. Diese anhand der Beobachtungsprotokolle vorzunehmen, muss als äußerst schwierig erachtet werden, da die Daten an sich bereits subjekt- und theoriegeleiteten Interpretationen durch die beobachtende Person unterliegen. Die genannten Schwächen der teilnehmenden Beobachtung, wie die Komplexität der Daten, deren begrenzte Dokumentierbarkeit und die Flüchtigkeit der Beobachtung, waren in Kombination mit den eingangs erwähnten Vorzügen der videobasierten Beobachtung die Hauptargumente für den Einsatz der Videografie als Datenerhebungsinstrument.

Mit der Entscheidung für den Einsatz der Videografie sind vor der eigentlichen Datenerhebung wichtige Vorbereitungen zu treffen. Zum einen müssen die Akteure im Feld ihre Zustimmung zur Teilnahme an der Videostudie geben. Zum anderen ist ein geeignetes Aufnahmekonzept, das eine Standardisierung der Aufnahmesituation ermöglicht, zu entwickeln und zu pilotieren. Wie diese beiden Schritte in der beschriebenen Dissertationsstudie bewältigt wurden, soll im Folgenden dargelegt werden.

Datenschutz

Bei Videodaten handelt es sich um äußerst sensible Daten. Anders als bei Audioaufnahmen oder Interviews ist eine komplette Anonymisierung der Daten nur unter großem technischen Aufwand möglich, was gleichzeitig mit einem erheblichen Authentizitätsverlust und einer Einschränkung der Aussagekraft des Datenmaterials einhergeht (vgl. Petko et al. 2003: 269). Insbesondere bei

Analysen, bei denen die Mimik der beobachteten Personen einbezogen werden soll, ist eine Anonymisierung der Daten problematisch. Dieser Umstand erfordert einen äußerst sensiblen und verantwortungsvollen Umgang mit dem erhobenen Videomaterial sowie den Richtlinien des Datenschutzes. Sollen Personen zu wissenschaftlichen Zwecken gefilmt werden, müssen alle an der Videostudie Beteiligten in schriftlicher Form ihr ausdrückliches Einverständnis dafür erteilen. Die rechtliche Grundlage für dieses Vorgehen findet sich im allgemeinen Persönlichkeitsrecht (APR) und ist im Grundgesetz rechtlich verankert (vgl. Bundeszentrale für politische Bildung 1998: 13).

Diese Einverständniserklärung, auch *informed consent form* (Mackey/Gass 2005: 26)[1], sollte einführende Informationen zur Durchführung und zum Ziel der Studie enthalten und transparent machen, wie die Daten bearbeitet, archiviert und anonymisiert werden. Darüber hinaus sollte eine Kontaktadresse für Rückfragen und im Falle von Dissertationsstudien auch die betreuenden Professoren bzw. Professorinnen genannt werden. Die angefragten Personen sollten in einem letzten Schritt der Erklärung unterzeichnen, dass sie a) über die Studie aufgeklärt worden sind b) damit einverstanden sind, dass von ihnen/ ihrem Kind Videoaufnahmen angefertigt werden und c) dass diese Daten zu wissenschaftlichen Zwecken verwendet werden dürfen.

In der vorliegenden Studie ist es durch den kontinuierlichen und intensiven Kontakt zum Forschungsfeld gelungen, eine hohe Kooperationsbereitschaft von Seiten der pädagogischen Fachkräfte und der Eltern (stellvertretend für ihre Kinder) zu erzielen[2]. Dazu trug auch ein möglichst hohes Maß an Transparenz bei, da bei Elternabenden, Themenabenden und in persönlichen Gesprächen mit den Erzieherinnen und Eltern offene Fragen geklärt und Zweifel größtenteils beseitigt werden konnten[3]. Dass diese hohe Teilnahmebereitschaft keine

---

1 Mackey/Gass (2005) diskutieren ausführlich die ethischen Gesichtspunkte, die im Vorfeld und während einer empirischen Studie mit Zweitsprachenlernenden bedacht werden sollten. In diesem Zusammenhang geben sie auch hilfreiche und praxisnahe Informationen zur Erstellung von Einverständniserklärungen und liefern Beispiele für den Aufbau derartiger Dokumente (ebd.: 33), die als Orientierungshilfe dienen können.

2 Die Kinder, für die keine von den Eltern unterschriebene Einverständniserklärung zur Teilnahme an der Videostudie vorlag, wurden an den Tagen der Videoaufzeichnungen in einer anderen Kita-Gruppe betreut.

3 Da die bilinguale Kita der TU Darmstadt von vielen Kindern mit Eltern nichtdeutscher Erstsprache besucht wird, wurden die Einverständniserklärungen in Anlehnung an die Vorschläge für die Erstellung von Einverständniserklärungen für nichterstsprachliche Proband/innen (vgl. Mackey/Gass 2005: 32) neben Deutsch

Selbstverständlichkeit ist, zeigen Angaben anderer videobasierter Studien wie der DESI-Studie, bei der lediglich 50% der angefragten Lehrkräfte bereit waren, ihren Unterricht videografieren zu lassen (vgl. u.a. Göbel 2010: 300). Die Teilnahmebereitschaft an Videostudien wird daher auch immer von der Gesamtgröße der Stichprobe beeinflusst. So war es in der beschriebenen Dissertationsstudie aufgrund der Stichprobengröße möglich, im Vorfeld einen persönlichen Kontakt zu allen potentiellen Teilnehmenden der Studie herzustellen, was sich vermutlich positiv auf die Teilnahmebereitschaft ausgewirkt hat. Dies ist jedoch bei Studien mit einer größeren Stichprobe oft nicht möglich.

Neben dem schriftlichen Einverständnis der pädagogischen Fachkräfte und der Eltern wurde auch die schriftliche Einwilligung des Trägers der Einrichtung eingeholt, Videoaufzeichnungen in den Räumlichkeiten der Kita anfertigen und diese zu wissenschaftlichen Zwecken nutzen zu dürfen. Wie in der TIMSS-Videostudie (vgl. Petko et al. 2003: 270) so unterzeichnete auch die wissenschaftliche Hilfskraft, die an der Datenerhebung und Aufbereitung der Daten beteiligt war, eine Vertraulichkeitsvereinbarung, in der sie sich verpflichtete, keinerlei personenbezogenen Informationen aus den Daten an Dritte weiterzugeben.

Es sei abschließend nochmals darauf hingewiesen, dass in Deutschland nach wie vor eine größtenteils ablehnende oder zumindest skeptische Haltung bezüglich der Teilnahme an Videostudien zu beobachten ist (vgl. Göbel 2010: 300). Gerade deshalb nimmt der Schritt der Anfrage bei den zu beobachtenden Personen und die Aufklärung über die Ziele und das Vorgehen der Studie in Form einer *informed consent form* eine Schlüsselposition in Videostudien ein.

Entwicklung eines geeigneten Aufnahmekonzepts

Im Rahmen wissenschaftlicher Analysen sollten Videodaten nach einem eigens entwickelten Aufnahmekonzept erhobenen werden, das eine Standardisierung und somit auch eine Qualitätssicherung und Vergleichbarkeit der Daten ermöglicht. Das Aufnahme- oder auch Kamerakonzept sollte dabei den spezifischen Charakteristika des Forschungsfeldes Rechnung tragen und relevante Daten zur Beantwortung der Forschungsfragen liefern. Für die großen Schulleistungsstudien wurden einige Aufnahmekonzepte erarbeitet und auch publiziert (siehe z.B. Petko 2006: 15-38). Wie bereits angedeutet, liegen für die Videografie von Bildungsprozessen im Elementarbereich wesentlich weniger und spärlicher methodische

---

auch auf Englisch ausgehändigt. Durch zusätzliche mündliche Rückfragen an die Eltern wurde sichergestellt, dass mindestens eines der beiden Elternteile den Inhalt der deutschen oder englischen Einverständniserklärung versteht und so eine informierte Einwilligung zum Filmen des eigenen Kindes geben kann.

Überlegungen vor (vgl. Ricart Brede 2011: 97). Bei der Entwicklung eines Aufnahmekonzepts für den Elementarbereich ist zu bedenken, dass schulische Lernsettings und der Kita-Alltag grundsätzlich verschieden sind. Während die beteiligten Akteure im Unterrichtsgeschehen zumeist feste Sitzplätze haben, ist die Kindergartenpraxis durch ein hohes Maß an Bewegungsaktivitäten charakterisiert. Zudem finden im Kindergarten viele Aktivitäten in einem Raum parallel statt, anstatt, wie im klassischen Unterrichtsgeschehen, auf die Person der Lehrkraft fokussiert zu sein. Eine nicht zu unterschätzende Herausforderung stellen des Weiteren die enorme Geräuschkulisse und der Lärmpegel in einer Kita dar. Insgesamt ist die Aufnahmesituation im Elementarbereich daher im Vergleich zur schulischen Unterrichtsbeobachtung als unübersichtlicher und komplexer zu charakterisieren.

In der vorliegenden Studie wurde die Aufnahmesituation Anfang 2012 pilotiert. In diesem Rahmen wurden an fünf Filmtagen verschiedene Kamerapositionierungen und Kameraeinsatzszenarien in einer der vier Kita-Gruppen erprobt. Die Vorüberlegungen zum Kameraeinsatz basieren dabei auf den Kameramanualen bestehender Videostudien im Elementarbereich (z.B. König 2006, Ricart Brede 2011) und beziehen sich im weiteren Sinne auch auf schulische Videostudien wie TIMSS und DESI. Mittels der Pilotierungsphase sollten in Vorbereitung auf die Hauptdatenerhebung insbesondere die Positionierung und Ausrichtung der Kamera(s) in den Kita-Gruppenräumen sowie die Bedienung der Kamera(s) getestet werden. Darüber hinaus sollten auch die Reaktion der beobachteten Personen auf die Kamerapräsenz und die Kameraeffekte in die Erarbeitung eines geeigneten Aufnahmekonzepts einfließen. Zudem wurde im Rahmen der Pilotierung der Frage nachgegangen, wie ein angemessenes Verhalten der Filmpersonen bei der Bedienung hinter der Kamera aussehen sollte, um das zu beobachtende Geschehen so wenig wie möglich zu beeinflussen.

Dieser Phase kommt im Forschungsprozess einer Videostudie eine tragende Rolle zu. Da die vorliegende Dissertationsstudie den Anspruch erhebt, natürliche und authentische Daten aus dem Alltag der bilingualen Kita zu erheben, die möglichst unbeeinflusst durch den Forscher zustande kommen (vgl. Silverman 2005), sollten die Aufzeichnungen ein Verhalten widerspiegeln, dass die Beobachteten auch ohne die Kameraanwesenheit zeigen würden. Gleichzeitig handelt es sich bei den videobasierten Beobachtungen dieser Studie um offene Beobachtungen, d.h. die Beobachteten wussten zu jedem Zeitpunkt der Aufnahme, dass sie aufgenommen werden. Allein dieses Wissen um die Anwesenheit der Kamera und die Aufnahmesituation kann allerdings auf Seiten der Beobachteten zu Verhaltensänderungen führen, die durch die Beobachtungssituation hervorgerufen werden. Dieses Phänomen wird in der Methodenliteratur zur Videografie als „Invasivität“ (Petko et al. 2003: 270) und zum Teil auch als „Reaktivität“ oder

„Reaktanz“ (Knoblauch/Schnettler 2007: 588) bezeichnet. Das Aufnahmekonzept dieser Studie sollte daher so geartet sein, dass mögliche Effekte der Erhebungssituation berücksichtigt und nach Möglichkeit Strategien zur Vermeidung von Invasivität eingesetzt werden.

Die Pilotierung hat zusammengefasst gezeigt, dass ein zurückhaltendes Verhalten seitens der Filmpersonen angebracht ist, um möglichst wenig Aufmerksamkeit auf die Erhebungssituation zu lenken. In Anlehnung an bereits bestehende Videostudien erwies es sich auch für dieses Vorhaben als vorteilhaft, sich als Filmperson vornehmlich auf die Bedienung der Kameraausrüstung zu konzentrieren und Interaktionen mit den Beobachteten und dem Forschungsfeld zu vermeiden (vgl. Petko et al. 2003: 270). Unterstützt werden sollte dies durch unauffällige und für Kinder uninteressante Kleidung, d.h. es sollte auf auffällige Farben und Geräusche durch Schmuck verzichtet werden. Zudem fanden die Aufnahmen in dem alltäglichen Gruppenraum der Kita statt, so dass sich sowohl die Kinder als auch die Erzieherinnen in ihrer gewohnten und vertrauten Umgebung befanden. Darüber hinaus erhielten die Erzieherinnen vor den Aufnahmen ein Schreiben zur Vorbereitung, in dem sie gebeten wurden, sich während der Aufnahmesituation möglichst normal zu verhalten und explizit gemacht wurde, dass an den Filmtagen keine speziell vorbereiteten und geplanten Aktivitäten stattfinden müssen. Um die Kinder besser auf die Aufnahmesituation vorzubereiten, wurden die Kameras bereits einige Tage vor der Pilotierung im Gruppenraum aufgestellt, so dass sich die Kinder an die Ausrüstung (Kamera, Stativ) gewöhnen konnten.

Um die Invasivität der Kamerapräsenz auf das Verhalten der Beobachteten abschätzen zu können, wurde im Anschluss an die Aufzeichnungen in der Pilotierung ein Reflexionsgespräch mit den pädagogischen Fachkräften geführt, in dem sie äußern durften, inwieweit sie sich von dem an diesem Tag erprobten Kamerakonzept „gestört“ und „beobachtet“ gefühlt haben. Da die Erzieherinnen die Kinder zudem besser kennen, konnten sie auch Auskunft darüber geben, wie die Aufnahmesituation das Verhalten der Kinder beeinflusst hat. Auf Grundlage dieser Rückmeldungen wurden die Aufnahmebedingungen stetig weiterentwickelt und die zu beobachtenden Personen in den Forschungsprozess und die Gestaltung der Aufnahmebedingungen aktiv mit einbezogen. Dies geschah unter der Annahme, dass die Personen nur dann ein natürliches Verhalten zeigen werden, wenn sie sich in der Aufnahmesituation nicht gestört, sondern wohl fühlen. Das Aufnahmekonzept, das schließlich der Hauptdatenerhebung zugrunde gelegt wurde, ist daher das Resultat eines gemeinsamen Entwicklungsprozesses, an dem sowohl die hauptverantwortliche Doktorandin und wissenschaftliche Hilfskraft als auch die Beobachteten, in diesem Fall die Erzieherinnen, beteiligt waren.

Die Erkenntnisse aus der Pilotierung und das entwickelte Aufnahmekonzept wurden schließlich in einem Kameraskript zusammengefasst. Dieses stellt im Rahmen von Videostudien ein wichtiges Instrument zur Standardisierung und Qualitätssicherung dar (vgl. Hall 2000). Durch das Aufstellen von Richtlinien zur Erhebungssituation (Position/Ausrichtung der Kameras, Bedienung der Kameras, Verhalten der Filmpersonen) ist es möglich, Videoaufnahmen untereinander vergleichbar zu machen (vgl. Petko 2006: 15). Mithilfe dieser ausführlichen Anleitung, die als Kameramanual oder Kameraskript bezeichnet wird, kann somit gewährleistet werden, dass die Erhebungsbedingungen an den verschiedenen Aufnahmetagen in der bilingualen Kita vergleichbar sind.

3.2 Datenerhebung

Nach der Pilotierung der Aufnahmebedingungen und einer erneuten Feinabstimmung der Erhebungsinstrumente fand die Hauptdatenerhebung in der bilingualen Kita der TU Darmstadt von Oktober bis Dezember 2012 statt. In diesem Zeitraum wurden alle vier Gruppen der Einrichtung an jeweils fünf Tagen videografiert. Es entstand somit ein Datenkorpus von über zwanzig Stunden Videomaterial. Während der Durchführung der videobasierten Datenerhebung zeigte sich, dass die Vorerfahrungen aus der Pilotierung insbesondere mit der Handhabung des Equipments eine unerlässliche Vorarbeit darstellten. So war die Aufgabenverteilung zwischen den durchführenden Personen (die hauptverantwortliche Doktorandin, die die Kamerabedienung übernahm und die wissenschaftliche Hilfskraft, die für die Überprüfung der Funktionstüchtigkeit vor jedem Aufnahmetag und den Aufbau der Kameraausrüstung in der Kita verantwortlich war) zu Beginn der Datenerhebung schon aufeinander abgestimmt und alle technischen Fragen im Rahmen der Pilotierung geklärt. In seinen Grundzügen konnten die im Kameraskript formulierten Richtlinien daher umgesetzt werden. Allerdings bringt es der Kita-Alltag mit sich, dass trotz sorgfältiger Vorbereitung nicht alle Vorkommnisse vorher planbar sind. Beispielsweise machten es Krankheitsfälle seitens der pädagogischen Fachkräfte oder die Eingewöhnungsphase von neuen Kindern in die Kita nötig, von den zuvor festgelegten Richtlinien abzuweichen, bzw. diese an den betreffenden Filmtagen flexibler zu handhaben und zu modifizieren. In diesem Zusammenhang weisen auch Petko et al. (2003: 271) darauf hin, dass eine restlose Standardisierung aufgrund unvorhersehbarer Zwischenfälle nicht möglich ist und das Kameraskript somit immer „einen Balanceakt zwischen Standardisierung und Offenheit“ darstellt. Um die Umsetzbarkeit und eventuelle Abweichungen vom Kameramanual zu dokumentieren, wurde begleitend zu den Filmtagen ein Aufnahmeprotokoll geführt, in

dem Aspekte der Umsetzbarkeit des Aufnahmekonzepts und auch Zwischenfälle bzw. Besonderheiten während der Aufnahmen protokolliert wurden. In Bezug auf die bereits genannten Invasivitätseffekte wurde auf einen anschließenden Fragebogen zur Typikalität der Aufzeichnungen, wie er in der TIMSS-Studie zum Einsatz gekommen ist (vgl. Petko et al. 2003: 270), bewusst verzichtet. Durch den intensiven Feldkontakt vor der Datenerhebung war es auch ohne nachträgliche Befragung möglich, die Typikalität der Videoaufzeichnungen einzuschätzen. Insgesamt kann wohl Petko et al. (vgl. 2003: 270) zugestimmt werden, wenn sie sagen, dass (Lehr-)Personen zwar tendenziell versuchen, einen aus ihrer Sicht von ihnen erwünschtes oder gutes Verhalten zu zeigen, doch muss auch angenommen werden, dass eine spontane und gleichzeitig grundlegende Verhaltensänderung aufgrund der Aufnahmesituation nicht ohne weiteres möglich sein wird (siehe z.B. auch Wahl 1991).

Nach Abschluss der Datenerhebung stellt die Aufbereitung und Auswertung der Videodaten den nächsten Arbeitsschritt dar. Im Folgenden soll daher ein Ausblick auf das geplante Vorgehen gegeben werden.

3.3 Ausblick auf Aufbereitung und Auswertung der Videodaten

Parallel zur Datenerhebung wurde mit der Aufbereitung der Videodaten begonnen. Diese umfasst die doppelte Speicherung der Daten an zwei verschiedenen Orten, um die Gefahr eines Datenverlusts zu verringern. Das Videomaterial, das in Full-HD-Qualität aufgenommen wurde, wurde in ein stark komprimiertes MPEG2-Format umgewandelt.[4] Die Komprimierung des Datenmaterials verfolgte dabei zwei Ziele. Zum einen sollte die zu speichernde Datenmenge reduziert werden (das MPEG2-Format ist weniger hochauflösend und benötigt daher weniger Speicherplatz). Zum anderen sollte es in ein Format umgewandelt werden, das von gängigen computerbasierten Analysetools unterstützt wird, da diese Analysesoftwares z.T. erhebliche Probleme mit zu großen Datenmengen haben. Das Ausgangsmaterial in Full-HD-Qualität steht jedoch nach wie vor zur Verfügung und kann beispielweise zu Präsentationszwecken zum Einsatz kommen. Zur Anonymisierung der Daten wurden den videografierten pädagogischen Fachkräften und den einzelnen Filmtagen Codes zugeordnet, die für Außenstehende eine Zuordnung zu Personen unmöglich machen.

Die Analyse der Videodaten soll in zwei aufeinander aufbauenden Schritten erfolgen. In einem ersten Schritt soll das Material makroanalytisch, d.h. auf der Ebene des gesamten videografierten Kita-Vormittags erfolgen, um in einem

4 Die Bearbeitung der Videodaten erfolgte mithilfe der Software *Pinnacle Studio 14.*

zweiten Schritt mikroanalytisch einzelne Sequenzen zu fokussieren. Dies ist ein gängiges Verfahren videogestützter Forschung und ein Forschungsdesign, das auch vielen bekannten Videostudien, wie beispielweise der DESI-Studie zugrunde liegt (vgl. Helmke et al. 2007: 38). Aufgrund der sequenziellen, aber auch logisch-inhaltlichen Verknüpfung dieser beiden Teilanalysen spricht man auch von einem *sequential mixed-method*-Stichprobengenerierungsverfahren (Teddlie/Yu 2007: 77), bei dem sowohl quantitative als auch qualitative Analyseverfahren zum Einsatz kommen. Die Annäherung an den Analysegegenstand mithilfe einer Kombination zweier oder mehrerer Methoden wird als *mixed methods approach* (vgl. Jick 1979) oder als Methoden-Triangulation bzw. methodologische Triangulation bezeichnet (siehe die Klassifikation nach Denzin (1970/1989), sowie die zusammenfassende Klassifikation im Einleitungsbeitrag von Settinieri in diesem Band). Der Einsatz mehrerer Methoden verfolgt dabei nicht das Ziel einer „einseitigen oder wechselseitigen Validierung der Einzelergebnisse" (Flick 2011: 19), sondern die eingesetzte Methodenvielfalt soll gemäß einer neueren Vorstellung von Triangulation zu einer umfassenden Gegenstandsabbildung beitragen (vgl. Flick 2011: 18). Vor allem, da es sich bei sprachlichen Interaktionsprozessen um einen äußerst komplexen Forschungsgegenstand handelt (vgl. Abschnitt 2 dieses Beitrags), bietet sich eine Kombination mehrerer Methoden im Analyseprozess an. Das Vorgehen der einzelnen Analyseschritte und die Kombination der Methoden im Sinne der Triangulation sollen im Folgenden erläutert werden.

Analyseschritte

In Vorbereitung auf die erste Teilanalyse sollen die Videodaten zunächst aufbereitet und in überschaubare Einheiten eingeteilt werden, welche die spätere Analyseeinheit bilden. Grundsätzlich stehen zur so genannten Sequenzierung, d.h. für die Einteilung des Materials in einzelne Sequenzen, zwei Verfahren zur Verfügung: das *time-sampling* oder das *event-sampling* (vgl. Bakeman/Gottman 1994). Beim *time-sampling* wird das gesamte Material in gleich lange Zeitintervalle, z.B. in Abschnitte von jeweils 30 Sekunden, unterteilt. Beim *event-sampling* hingegen erfolgt die Sequenzierung ereignisbezogen, d.h. immer wenn eine neue Aktivität einsetzt, wird ein Sequenzwechsel kodiert. So entstehen durch die inhaltsbezogene Segmentierung[5] Handlungsphasen und Videoabschnitte unterschiedlicher Länge (vgl. Petko et al. 2003: 274). Das *event-sampling* ist ein durchaus übliches Verfahren der videogestützten Forschung, das auch in anderen

5 Dinkelaker/Herrle (2009: 52) bezeichnen dieses Vorgehen daher auch als Segmentierungsanalyse.

Studien zum Einsatz gekommen ist (vgl. z.B. Hugener et al. 2006: 48). Auf der Grundlage dieser Sequenzierung soll dann im Rahmen der ersten Teilanalyse eine Basiskodierung mithilfe eines niedrig-inferenten Beobachtungsinstruments vorgenommen werden. Bei niedrig-inferenten Beobachtungsinstrumenten handelt es sich um Kategorien, die direkt beobachtbar sind, d.h. auf Sichtstrukturen beruhen, während hoch-inferente Beobachtungsinstrumente ein höheres Maß an interpretativen Schlussfolgerungen (Inferenzen) erfordern (vgl. Clausen et al. 2003, Seidel et al. 2003). Mithilfe der niedrig-inferenten Basiskodierung soll daher eine möglichst präzise Beschreibung der inhaltlichen Nutzung der Freispielzeit, d.h. der Gestaltung, der Ereignisse und Organisationsformen vorgenommen werden. Als Ziel und Ergebnis dieses Schritts steht daher die Identifizierung von Handlungssequenzen und inhaltlichen Gestaltungsmerkmalen, die die rahmengebenden Aktivitäten für das Zustandekommen von sprachlichen Interaktionen zwischen pädagogischen Fachkräften und Kindern während der Freispielzeit in der bilingualen Kita (vgl. erste Teilfrage der Forschungsfrage) bilden.

Die Diskussion der Methode der teilnehmenden Beobachtung hat bereits gezeigt, dass Beobachtungen anfällig für subjektive Einflüsse durch den Beobachter sind und durch theorie- und vorwissensgeleitete Interpretationen eingefärbt sein können. Um „Verzerrungen durch die Person des Forschers aufzudecken bzw. zu minimieren“ (Flick 2011: 14) und zu prüfen, ob die entwickelten Beobachtungsinstrumente transparent und intersubjektiv nachvollziehbar sind, wird die Basiskodierung für einen Teil des Datensatzes durch zwei verschiedene Beobachter/innen vorgenommen. Der Einsatz von zwei Beobachter/innen wird in Anlehnung an die Klassifikation von Triangulations-Typen nach Denzin (1970/1989) als *Investigator Triangulation* oder auch Beobachtertriangulation bezeichnet (Denzin 1970/1989, zitiert nach Flick 2011: 12f.).

Auf Grundlage der eher quantitativ angelegten niedrig-inferenten Basiskodierung sollen im zweiten Analyseschritt aus dem Pool der beschriebenen Handlungseinheiten einzelne Sequenzen ausgewählt und mikroanalytisch untersucht werden. Um die Charakteristika der sprachlichen Interaktion aus der Perspektive der pädagogischen Fachkraft präzise beschreiben zu können, müssen die Videosequenzen transkribiert und in einzelne Turns segmentiert werden. Anhand der erstellten Transkripte soll dann untersucht werden, wie Erzieherinnen ihr sprachliches Verhalten in der Interaktion mit den ein- bis dreijährigen Kindern gestalten und welche Charakteristika diesen Interaktionsprozessen zugrunde liegen. Insbesondere soll in diesem zweiten Analyseschritt auf Aspekte der Responsivität und den Einsatz von *Scaffolding*-Elementen (u.a. Kontextualisierungsstrategien und Strategien der Verständnissicherung) eingegangen werden. Für diesen Auswertungsschritt soll ein Beobachtungssystem entwickelt werden, das

die charakteristischen *Scaffolding*-Elemente in der bilingualen Kita möglichst gegenstandsangemessen abbildet. Dazu sollen zunächst aus der Forschungsliteratur und bereits bestehenden Studien vorläufige Kategorien deduktiv in das Beobachtungssystem aufgenommen werden, um diese dann am Material zu pilotieren und um weitere, induktiv aus dem Material ableitbare Kategorien zu erweitern und auszudifferenzieren.

Bezug nehmend auf die zweite Teilfrage der Forschungsfrage wird dieser qualitative und mikroanalytische Auswertungsschritt anhand ausgewählter Beispiele aus der Praxis der bilingualen Kita Aufschluss darüber geben, wie sprachliche Interaktionsprozesse zwischen pädagogischen Fachkräften und Kleinkindern in unterschiedlichen Situationen organisiert sind und welche Merkmale das sprachliche Interaktionsverhalten der deutsch- und englischsprachigen Fachkräfte gegenüber den Kindern kennzeichnen.

4. Fazit und Ausblick

Anliegen des vorliegenden Beitrags war es, die Einsatzmöglichkeiten der Methode der Videografie im Elementarbereich am Beispiel einer Dissertationsstudie zur Erforschung von sprachlichen Interaktionsprozessen in einer bilingualen Kita aufzuzeigen. Die einleitenden Ausführungen haben verdeutlicht, dass die videobasierte Unterrichtsforschung vor allem vor dem Hintergrund technischer Neuerungen wie der Digitalisierung und der größeren Speicherkapazität von Computern und angesichts der Grenzen der teilnehmenden Beobachtung einen enormen Entwicklungsschub erfahren hat. So ist es heute im Gegensatz zu noch vor einigen Jahren aus technischer Sicht mit einem vergleichsweise geringen Aufwand möglich, große Mengen an audiovisuellen Daten zu produzieren (vgl. Abschnitt 1). Die methodischen Überlegungen im Kontext der beschriebenen Studie sollten jedoch auch veranschaulichen, dass neben technischen Gesichtspunkten noch weitaus mehr Faktoren bei der Entwicklung eines videobasierten Forschungsdesigns und der Vorbereitung einer Datenerhebung mittels Videgrafie berücksichtigt werden müssen. Dies betrifft vor allem die Überlegungen zum Datenschutz, die in Videostudien eine nicht zu unterschätzende Bedeutung haben sowie die Kamera- und Invasivitätseffekte, die bei der Entwicklung eines Aufnahmekonzepts bedacht werden sollten (vgl. Abschnitt 3).

Im Zuge videogestützter Beobachtungsverfahren entstehen komplexe und reichhaltige Daten. Aus diesem Grund sollte die methodische Herangehensweise an Videostudien ebenso mehrperspektivisch erfolgen. Die Videografie bietet sich daher für die Einnahme von unterschiedlichen Perspektiven auf einen Forschungsgegenstand und die Beantwortung von Forschungsfragen im Sinne

der Triangulation geradezu an. Das beschriebene Forschungsdesign versucht durch eine Kombination unterschiedlicher Methoden (Methodentriangulation) mit einer ersten quantifizierenden Basiskodierung, auf die die qualitative Mikroanalyse einzelner Sequenzen folgt, einen prinzipiellen Erkenntniszuwachs anzustreben (vgl. Flick 2011: 12). Ebenso verfolgt der Einsatz von zwei Beobachter/innen (Beobachtertriangulation/Forschertriangulation) bei der Basiskodierung das Ziel, Beobachtereinflüsse abzufedern und zu vertieften Erkenntnissen zu gelangen. Exemplarisch sei hier auch auf den vielversprechenden Vorschlag von Schramm/Aguado (vgl. 2010: 198) hingewiesen, die deskriptive Analyse von Videodaten mit introspektiven Datenquellen (z.B. lautes Denken, video-stimulierte retrospektive Interviews, videostimulierte Retrospektion) im Sinne einer Datentriangulation zu kombinieren.

Da die technischen Herausforderungen der Videografie mittlerweile bewältigt sind, sollte sich die zukünftige Diskussion um den Einsatz der Videografie verstärkt der methodologischen Weiterentwicklung des Instruments und der Systematisierung von Auswertungsansätzen videobasierter Beobachtungsdaten widmen. Dieser Punkt scheint angesichts der Publikationslage bis dato eine Lücke darzustellen. In diesem Kontext sollten vermehrt auch Ansätze der Triangulation in die Diskussion einbezogen werden, da es insbesondere für Videostudien lohnenswert sein kann, durch eine Kombination von verschiedenen Methoden, Daten, Theorien oder den Einsatz mehrerer Beobachter/innen einen mehrperspektivischen Zugang zum Forschungsgegenstand einzunehmen und dadurch einen prinzipiellen Erkenntniszuwachs bzw. Erkenntnisse auf unterschiedlichen Ebenen zu erzielen.

Literatur

Aguado, Karin / Schramm, Karen / Vollmer, Helmut Johannes (Hrsg.) (2010). *Fremdsprachliches Handeln beobachten, messen, evaluieren*. Reihe KFU, Band 37. Frankfurt/Main: Lang.

Aufschnaiter, Stefan von / Welzel, Manuela (Hrsg.) (2001). *Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen. Aktuelle Methoden empirischer pädagogischer Forschung*. Münster: Waxmann.

Bakeman, Roger / Gottman, John (1994). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Bortz, Jürgen / Döring, Nicola (2006). *Forschungsmethoden und Evaluation*. 4. Aufl. Heidelberg: Springer.

Bundeszentrale für politische Bildung (Hrsg.) (1998). *Grundgesetz für die Bundesrepublik Deutschland*. Ulm: Ebner.

Clausen, Marten / Reusser, Kurt / Klieme, Eckhard (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft* 31/2, 122-141.

Denzin, Norman ([1970] 1989). *The Research Act*. 3. Aufl. Englewood Cliffs, N.J.: Prentice Hall.

Dinkelaker, Jörg / Herrle, Matthias (2009). *Erziehungswissenschaftliche Videografie. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Döpke, Susanne (1992). *One Parent-One Language. An Interactional Approach*. Amsterdam: Benjamins.

Europäische Kommission (1996). *Weißbuch Lehren und Lernen: Auf dem Weg zur kognitiven Gesellschaft*. Online. http://ec.europa.eu/languages/documents/doc409_de.pdf [10.3.2014].

Fischer, Dietlind / Schratz, Michael (2005). Videos in der Lehrerbildung. *Journal für LehrerInnenbildung* 5/2, 4-7.

Flick, Uwe (2011). *Triangulation. Eine Einführung*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick, Uwe (2004). *Qualitative Sozialforschung. Eine Einführung*. Reinbek: Rowohlt.

FMKS (2010). *Bilinguale Kitas in Bundesländern*. Stand August 2010. Online: http://www.fmks-online.de/download.html# [10.3.2014].

Genesee, Fred (1987). *Learning through Two Languages: Studies of Immersion and Bilingual Education*. Cambridge, MA: Newbury House.

Göbel, Kerstin (2010). Videografie als Verfahren zur Erforschung von Lehr-/Lernprozessen im Fremdsprachenunterricht. In: Aguado, Karin et al. (Hrsg.), 277-305.

Hall, Rogers (2000). Videorecording as Theory. In: Kelly, Anthony / Lesh, Richard (Hrsg.). *Handbook of Research Design in Mathematics and Science Education*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 647-664.

Helmke, Andreas / Helmke, Tuyet / Kleinbub, Iris / Nordheider, Iris / Schrader, Friedrich-Wilhelm/ Wagner, Wolfgang (2007). Die DESI-Videostudie. Unterrichtstranskripte für die Lehrerausbildung nutzen. *Der fremdsprachliche Unterricht Englisch* 90, 37-45.

Hugener, Isabelle / Rakoczy, Katrin / Pauli, Christine / Reusser, Kurt (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In: Rahm, Sibylle/ Mammes, Ingelore/ Schratz, Michael (Hrsg.). *Schulpädagogische*

Forschung, Band 1, Unterrichtsforschung – Perspektiven innovativer Ansätze. Innsbruck: Studienverlag, 41-53.

Jick, Todd D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly* 24/4, 602-611.

Knoblauch, Hubert / Schnettler, Bernt (2007). Videografie. Erhebung und Analyse qualitativer Videodaten. In: Buber, Renate / Holzmüller, Hartmut (Hrsg.). *Qualitative Marktforschung. Theorie, Methode, Analysen*. Wiesbaden: Gabler, 585-599.

König, Anke (2006). *Dialogisch-entwickelnde Interaktionsprozesse zwischen ErzieherIn und Kind(-ern). Eine Videostudie aus dem Alltag des Kindergartens. Inauguraldissertation*. Dortmund. Online: https://eldorado.tu-dortmund.de/bitstream/2003/24563/1/Diss_veroeff.pdf [10.3.2014].

Long, Michael (1996). The role of linguistic environment in second language acquisition. In: Ritchie, William / Bhatia, Tej (Hrsg.). *Handbook of Second Language Acquisition*. San Diego: Academic Press, 413-468.

Mackey, Alison / Gass, Susan M. (2005). *Second Language Research. Methodology and Design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Pauli, Christine / Reusser, Kurt (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik* 52/6, 774-798.

Petko, Dominik / Waldis, Monika / Oauli, Christine / Reusser, Kurt (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. Ansätze der TIMSS 1999 Video Studie und ihrer schweizerischen Erweiterung. *ZDM (Zentralblatt für Didaktik der Mathematik)* 35/6, 265-280. Online: http://subs.emis.de/journals/ZDM/zdm036a4.pdf [10.3.2014].

Petko, Dominik (2006). Kameraskript. In: Klieme, Eckhard / Pauli, Christine / Reusser, Kurt (Hrsg.). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“. Teil 3: Videoanalysen*. Frankfurt/Main: GFPF, 15-38.

Ricart Brede, Julia / Knapp, Werner / Gasteiger Klicpera, Barbara / Kucharz, Dietmut (2010). Die Entwicklung von Beobachtungssystemen in der videobasierten Forschung am Beispiel videobasierter Qualitätsanalysen vorschulischer Sprachfördereinheiten. In: Aguado, Karin et al. (Hrsg.), 257-275.

Ricart Brede, Julia (2011). *Videobasierte Qualitätsanalyse vorschulischer Sprachfördereinheiten*. Freiburg im Breisgau: Fillibach.

Schramm, Karen / Aguado, Karin (2010). Videografie in den Fachdidaktiken - Ein Überblick. In: Aguado, Karin et al. (Hrsg.), 185-214.

Seidel, Tina / Kobarg, Mareike / Rimmele, Rolf (2003). Aufbereitung der Videodaten. In: Seidel, Tina / Prenzel, Manfred / Duit, Reinders / Lehrke, Manfred (Hrsg.). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht"*. Kiel: IPN, 47-67.

Settinieri, Julia (2014). Forschst Du noch, oder triangulierst Du schon? In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 17-35.

Silverman, David (2005). *Doing Qualitative Research. A Practical Handbook*. London: SAGE Publications.

Teddlie, Charles / Yu, Fen (2007). Mixed Methods Sampling. A Typology With Examples. *Journal of Mixed Methods Research* 1/1, 77-100.

Wahl, Diethelm (1991). *Handeln unter Druck. Der weite Weg vom Wissen zum Handeln bei Lehrern, Hochschullehrern und Erwachsenenbildnern*. Weinheim: Deutscher Studien Verlag.

Wode, Henning (1995). *Lernen in der Fremdsprache: Grundzüge von Immersion und bilingualem Unterricht*. Ismaning: Hueber.

Wygotski, Lev (1987). *Ausgewählte Schriften. Band 2: Arbeiten zur psychischen Entwicklung der Persönlichkeit*. Köln: Pahl-Rugenstein.

Zydatiß, Wolfgang (2000). *Bilingualer Unterricht in der Grundschule: Entwurf eines Spracherwerbskonzepts für zweisprachige Immersionsprogramme*. Ismaning: Hueber.

Ansätze zur Triangulation mithilfe qualitativer Forschungssoftware (MAXQDA 11)

Claus Stefer

Dieser Beitrag fokussiert den Einsatz von MAXQDA 11, einer Software zur Analyse qualitativer Daten, im Kontext von triangulativen Datenauswertungsverfahren. Es werden die häufigsten und flexibelsten Triangulationsstrategien anwendungsorientiert aufgegriffen. Dargestellt werden sowohl einfache häufigkeitsorientierte Verfahrensweisen als auch komplexere Techniken wie die Arbeit mit Variablen, Joint Displays, Summary Grids, Elemente der quantitativen Inhaltsanalyse sowie die Intercoder-Übereinstimmug als grundlegender Ansatz der Forscher-/innentriangulation. Der Beitrag ist auf die Forschungspraxis hin ausgerichtet, d.h., dass die einzelnen Vorgehensweisen und die dahinterliegenden methodologischen Annahmen zunächst knapp vorgestellt werden, um im Anschluss ihre Umsetzung mithilfe von MAXQDA in den Blick zu nehmen.

1. Einleitung

Software zur Unterstützung des Forschungsprozesses hat sich als selbstverständlicher Teil moderner sozialwissenschaftlicher Forschung etabliert. Verfolgte diese Art Software zunächst vor allem das Ziel, etablierte Arbeitsabläufe qualitativer Forschung auf den Computer zu übertragen und so die Arbeit übersichtlicher und effizienter zu gestalten, so hat sich im Laufe von nunmehr über zwei Dekaden eine eigene Familie von Software für die qualitative Datenanalyse entwickelt, die durch das breite Spektrum angebotener Analysemöglichkeiten neue Perspektiven für die Datenauswertung eröffnet. Diese Art Analyseprogramme wird im deutschen Raum in der Regel als *QDA*-Software bezeichnet – das Kürzel steht für *Qualitative Datenanalyse* –, im englischsprachigen Raum ist die Bezeichnung CAQDAS (*Computer Assisted Qualitative Data Analysis Software*) gebräuchlich. Während der fortlaufenden Entwicklung konnte die Funktionalität der jeweiligen Programme nicht zuletzt aufgrund der rasant wachsenden Leistungsfähigkeit von Computern stetig wachsen. Die letzte Generation von QDA-Programmen kann demnach mit einer Fülle von Funktionen und Möglichkeiten aufwarten, die vor wenigen Jahren noch undenkbar gewesen wären. Sie eröffnen völlig neue Perspektiven auf die Datenanalyse und können für eine Vielzahl qualitativer Forschungsvorhaben eingesetzt werden. Mit ihnen lassen sich enorme Datenmengen verwalten und bearbeiten, sie erlauben die Integration

unterschiedlichster Dokument- und Medienarten, bieten flexible Teamwork-Funktionen an und erhöhen die Nachvollziehbarkeit des Forschungsprozesses. Eine besondere Stärke der Software liegt in ihrer Fähigkeit, unterschiedliche Forschungsmethoden und -logiken zu integrieren, um sie möglichst selbstverständlich auf denselben Datenkorpus anwenden zu können. Aber auch die Zusammenführung von Daten aus unterschiedlichen methodischen Quellen bzw. die Bereitstellung der Arbeitsergebnisse für die weitere Analyse mithilfe anderer Software beherrscht moderne QDA-Software. Somit bietet sie optimale Voraussetzungen, die mehr und mehr üblichen Wege der *Triangulation*, *Methodenintegration* und *Mixed Methods* effizient in der Forschungspraxis umzusetzen[1]. Dieser Beitrag untersucht anhand einer Betrachtung der Software *MAXQDA* (Version 11), welche konkreten Möglichkeiten zur Triangulation QDA-Software anbietet.

2. Grundkonzepte der Triangulation in QDA-Software

QDA-Software stellt verschiedene Techniken zur Verfügung, Triangulationen vorzunehmen. Folgende Verfahrensweisen stellen dabei die dominanten dar:

Integration qualitativer und quantitativer Daten und Forschungslogiken: Dieser Aspekt beschreibt die Möglichkeit, standardisierte und nicht standardisierte Daten in einem einzigen Datensatz zusammenzubringen, und zwar in einer Weise, die es einerseits ermöglicht, die Daten jeweils separat zu analysieren, sie jedoch auch unmittelbar und direkt miteinander zu verbinden. Ein Weg kann sein, separat vorliegende Daten in einer Software zusammenzuführen, um so alle Informationen unmittelbar kombiniert vorliegen zu haben, sie in der Analyse aufeinander beziehen zu können und so anzureichern. Einen anderen Weg stellt die Möglichkeit dar, Forschungslogiken zu kombinieren, etwa direkt bei der interpretativ-bewertenden Einordnung von Textstellen eine Likert-Skala

1 Dieser Beitrag befasst sich nicht mit den Unterschieden zwischen den Konzepten der *Triangulation, Methodenintegration* und der *mixed methods,* die jeweils eigene Strategien und Vorgehensweisen verfolgen. Stattdessen werden diese Begriffe im vorliegenden Beitrag auf einer sehr allgemeinen Ebene synonym verwendet. Damit wird vor allem der diesen Ansätzen gemeinsame Aspekt betont, Daten anzureichern, zu validieren, zu ergänzen und die Aussagekraft und Qualität einer Untersuchung zu steigern. Unter dem Begriff *Triangulation* werden hier somit allgemein Bestrebungen zusammengefasst, numerische und nicht numerische, standardisierte und nicht standardisierte Daten gemeinsam auszuwerten, aufeinander zu beziehen oder in die jeweils andere Form zu transformieren. Diese weit gefasste Begriffsverwendung soll keineswegs eine Gleichheit der verschiedenen Konzepte zum Ausdruck bringen. Für eine genauere Ausdifferenzierung der Begriffe vgl. Settinieri (in diesem Band).

anzuwenden. Darüber hinaus existieren Schnittstellen, die es ermöglichen, die Daten und Inhalte, die in der QDA-Software bearbeitet werden, auch an Statistik-Software zu übergeben, um speziellere Analysen durchführen zu können.

Quantifizierung der qualitativen Analyse: QDA-Software ermöglicht es, verschiedenste Schritte, Tätigkeiten und Facetten qualitativer Analysearbeit auch in quantifizierter Form darzustellen. Wichtige Werkzeuge können hier etwa Codehäufigkeiten, die Anzahl der Überschneidungen von bestimmten Codes oder auch der prozentuale Anteil eines Dokumentes, der codiert wurde, (Codierdichte) sein. Auch diese Informationen können an Statistik-Software übergeben werden, etwa um weitergehende Zusammenhangsanalysen durchführen zu können.

Joint Displays: Sie ermöglichen, qualitative und quantitative Informationen in einer einzigen Übersicht zu vereinen und so gemeinsam darzustellen. Sie erlauben beispielsweise, die Aussagen nach quantitativen Merkmalen und Eigenschaften sortiert auszugeben und gegenüberzustellen, etwa die Aussagen der befragten Personen sortiert nach ihrem höchsten formalen Bildungsabschluss.

Quantitative Analyse qualitativen Materials: Neben rein oder vornehmlich interpretativen Vorgehensweisen existiert auch eine Reihe von Verfahren, mit deren Hilfe nicht standardisiertes – also qualitatives – Datenmaterial auf statistischer Basis ausgewertet werden kann, so etwa die *Quantitative Inhaltsanalyse*. Auch die vergleichsweise einfache Analyse von Worthäufigkeiten, die einen spannenden Einblick in den Wortschatz der zu analysierenden Dokumente und somit auch in die ausgedrückte Weltsicht des Dokumentes ermöglicht, ist dieser Art der Triangulation zuzurechnen.

Teamwork: Moderne QDA-Software ermöglicht auf vielfältige Weise die Kooperation von Forschenden und somit die Forscher-/innentriangulation. Aber auch die Untersuchung von Intercoder-Übereinstimmungen, die ein Merkmal für die Zuverlässigkeit der vorgenommenen Codierungen darstellt, wird ohne Weiteres ermöglicht.

Auf diese fünf Grundkonzepte aufbauend werden in QDA-Software unterschiedlichste Analysekonzepte umgesetzt, die es den Forschenden ermöglichen, Triangulationen den spezifischen Anforderungen entsprechend vorzunehmen.

3. Triangulation in MAXQDA 11

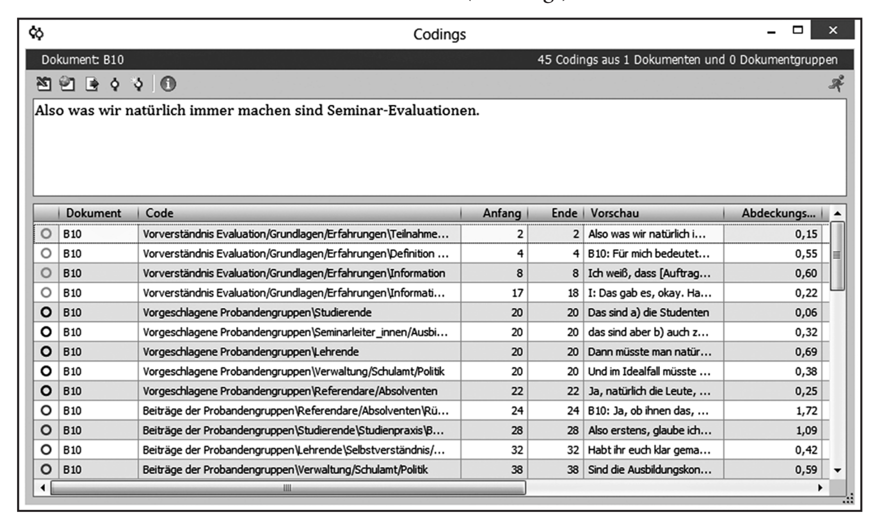
In MAXQDA 11 stehen Forschenden vielfältige Möglichkeiten zur Triangulation zur Verfügung. Nachfolgend werden die zentralen Funktionen der Software betrachtet. Dabei können an dieser Stelle nicht alle Funktionen der Software im Detail besprochen werden, sondern es wurde eine Auswahl getroffen, die sich am zu erwartenden Nutzwert der Funktionen für ein breites Publikum orientiert und für eine große Zahl unterschiedlicher Forschungskontexte nutzbar ist.

Vor der Betrachtung der einzelnen Funktionen ist ein prinzipieller Hinweis auf einen aus qualitativ-interpretativer Perspektive wesentlichen Aspekt der Arbeit mit Quantifizierungen unerlässlich: Quantitäten, beispielsweise die Häufigkeit von Codierungen, sind zwar als unzweifelhaftes Faktum zu betrachten, ein direkter Rückschluss von Häufigkeiten auf inhaltliche Relevanzen ist allerdings nicht unbedingt zulässig. Nur weil etwa eine Kategorie während des Codierens besonders häufig angewendet wurde, ist sie nicht automatisch besonders wichtig. Genauso gut könnte sie einfach die ungenaueste Kategorie sein oder aber diejenige, die bewusst am breitesten definiert wurde. Andererseits macht es einen deutlichen Unterschied, ob beispielsweise ein Code, der qua Definition in jedem Dokument maximal einmal vergeben werden kann, bei einem 30 Dokumente umfassenden Forschungsprojekt vier- oder 24-mal angewendet wurde. Die Forscher/innen haben demnach stets die Aufgabe, die Erkenntnisse oder Vermutungen, die sich aus der Arbeit mit Quantifizierungen ergeben, an den qualitativen Kontext rückzubinden, aus dem sie hervorgehen bzw. den sie abbilden, sie daran zu überprüfen und die inhaltliche, qualitative Dimension dieser Zahlenwerte zu erschließen. Die substanzielle Zusammenführung und Rückbindung der einzelnen Elemente ist somit von entscheidender Bedeutung. Oder anders formuliert: Die Zahlen lügen nicht, bedürfen aber spezifischer Analyse.

3.1 Grundlegende Quantifizierungen: Code- und Überschneidungshäufigkeiten

Die grundlegendste Stufe der Quantifizierung steht in MAXQDA permanent und ohne gesonderten Aufruf zur Verfügung: Hinter jedem Dokument wird die Anzahl der in ihm codierten Dokumentstellen in Zahlenform angezeigt. Gleiches gilt für die Codes, auch hier wird hinter jedem Code als Zahl ausgegeben, wie viele Dokumentstellen ihm zugeordnet wurden. Diese Zahlen werden automatisch während der Arbeit mit den Daten erzeugt und aktualisiert und bilden somit die interpretative Arbeit numerisch ab. Allerdings bleiben diese Werte immer auf Ebene eines Dokumentes oder eines Codes und sind kumulativ, also nicht differenziert. Um diese Werte differenzierter zu betrachten, kann mithilfe der Funktion *Übersicht Codings* sowohl für ein Dokument als auch für einen Code eine tabellarische Übersicht aller im Dokument codierten bzw. dem Code zugeordneten Passagen aufgerufen werden. Dazu wird das betreffende Dokument oder der betreffende Code mit der rechten Maustaste angeklickt, um aus dem erscheinenden Menü den Eintrag *Übersicht Codings* aufzurufen.

Abb. 1: Übersicht über alle codierten Textstellen (=Codings) des Dokuments B10.



Dokument	Code	Anfang	Ende	Vorschau	Abdeckungs...
B10	Vorverständnis Evaluation/Grundlagen/Erfahrungen\Teilnahme...	2	2	Also was wir natürlich i...	0,15
B10	Vorverständnis Evaluation/Grundlagen/Erfahrungen\Definition ...	4	4	B10: Für mich bedeutet...	0,55
B10	Vorverständnis Evaluation/Grundlagen/Erfahrungen\Information	8	8	Ich weiß, dass [Auftrag...	0,60
B10	Vorverständnis Evaluation/Grundlagen/Erfahrungen\Informati...	17	18	I: Das gab es, okay. Ha...	0,22
B10	Vorgeschlagene Probandengruppen\Studierende	20	20	Das sind a) die Studenten	0,06
B10	Vorgeschlagene Probandengruppen\Seminarleiter_innen/Ausbi...	20	20	das sind aber b) auch z...	0,32
B10	Vorgeschlagene Probandengruppen\Lehrende	20	20	Dann müsste man natür...	0,69
B10	Vorgeschlagene Probandengruppen\Verwaltung/Schulamt/Politik	20	20	Und im Idealfall müsste ...	0,38
B10	Vorgeschlagene Probandengruppen\Referendare/Absolventen	22	22	Ja, natürlich die Leute, ...	0,25
B10	Beiträge der Probandengruppen\Referendare/Absolventen\Rü...	24	24	B10: Ja, ob ihnen das, ...	1,72
B10	Beiträge der Probandengruppen\Studierende\Studienpraxis\B...	28	28	Also erstens, glaube ich...	1,09
B10	Beiträge der Probandengruppen\Lehrende\Selbstverständnis/...	32	32	Habt ihr euch klar gema...	0,42
B10	Beiträge der Probandengruppen\Verwaltung/Schulamt/Politik	38	38	Sind die Ausbildungskon...	0,59

Eine weitergehende, vergleichende und daher besser für das Erlangen eines Überblicks geeignete Darstellung der Codehäufigkeiten stellt der *Code-Matrix-Browser* (Menü *Visual Tools*) zur Verfügung. Er erlaubt, die Häufigkeiten der Codierungen für jeden Code und jedes Dokument gleichzeitig kumuliert in grafischer Form in einer Tabelle darzustellen. Der Code-Matrix-Browser ist ein Gitter, in dem jede vorgenommene Codierung in Form eines Quadrates repräsentiert wird. Je häufiger ein bestimmter Code in einem Dokument vergeben wurde, desto größer wird das Quadrat in der Grafik dargestellt. Wurde ein Code nicht angewendet, bleibt die betreffende Position leer. Neben dieser grafischen Form der Darstellung ermöglicht der Code-Matrix-Browser aber auch, die absoluten Codierhäufigkeiten als Zahlwerte auszugeben. Diese alternative Anzeigeform ist deutlich präziser, da die grafischen Symbole kleine Unterschiede der Häufigkeiten unter Umständen nicht abbilden können, dafür sind Maxima und Minima schwieriger auszumachen. Die Informationen dieser Grafik, also die Anzahl der codierten Textstellen je Code in jedem Dokument, lassen sich ohne Weiteres über ein entsprechendes Icon in Form einer Tabelle exportieren. Dadurch besteht eine direkte Anknüpfung an die Welt der statistischen Datenanalyse, beispielsweise um mithilfe einer Software für statistische Analysen Korrelationen zwischen Codes zu berechnen.

Abb. 2: Ausschnitt aus einem Code-Matrix-Browser.

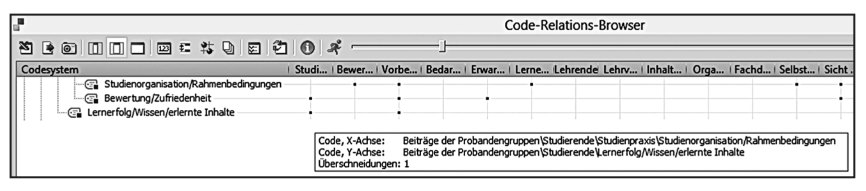
Code-Matrix-Browser

Codesystem | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | B10

Lernstrategien
Individuelle Voraussetzunge...
Lernverlauf
Teilnahmeverlauf an de...
die alles Wahrnehm...
VL und Übungsbasiert
Tutoriumslerner
Lernen außerhalb der V...
lerne nicht
lerne viel
Erwartung Klausurnote
gut
zwei bis drei
mittel
Hauptsache durch

Als Ergänzung zu dieser Übersichtsdarstellung der Codehäufigkeiten erlaubt der *Code-Relations-Browser* (Menü *Visual Tools*) in gleicher Darstellungsform, also sowohl grafisch als auch numerisch, die Anzeige der Überschneidungen und Nähe von Codes. Die erzeugte Grafik gibt folglich Aufschluss darüber, wie vielen Textstellen mindestens zwei Codes zugewiesen wurden. Dazu wird eine Matrix erzeugt, die sowohl in den Spalten als auch in den Zeilen frei wählbare Codes enthält. An der Schnittstelle von Zeilen und Spalten wird jeweils als Grafik oder Zahl ausgegeben, wie häufig sich die beiden betreffenden Codes überschneiden. Dabei kann definiert werden, ob es sich um reale Überschneidungen handelt, eine Textstelle also tatsächlich beiden Codes zugewiesen wurde, oder ob sich die Codierungen in der Nähe zu einander befinden. Die Nähe wird dabei in Form von Absätzen definiert, die zwischen dem Vorkommen der beiden Codes im gleichen Text liegen können.

Abb. 3: Suche nach tatsächlichen Überschneidungen der drei Codes in der ersten Spalte mit anderen Codes anhand der Darstellung im Code-Relations-Browsers.



Diese Darstellungen der Quantitäten gestatten den Forschenden, einen schnellen Überblick über die Vergabe von Codes zu gewinnen und so einen neuen, ergänzenden Eindruck von der zahlenmäßigen Bedeutung des Kategoriensystems und seiner Teile zu erhalten. Ein Blick auf die bloßen Häufigkeiten der Codierungen je Code und je Dokument sowie auf die jeweiligen prozentualen Anteile kann Hinweise bezüglich der Relevanz einer Kategorie, der Erfordernis einer neu vorzunehmenden,

präziseren Codedefinition oder der inhaltlichen Dichte eines Dokumentes hervorbringen. Dem Code-Matrix-Browser kann unter anderem schnell entnommen werden, welche Dokumente besonders viele Codierungen zu einem bestimmten Code enthalten oder welcher Code in einem Dokument der dominante ist, und zwar jeweils, wenn gewünscht, im Kontext aller bearbeiteten Dokumente. Der Code-Relations-Browser zeigt auf, welche Codes häufig gleichzeitig vergeben wurden, somit eventuell von ihrer inhaltlichen Bedeutung her zusammentreffen, oder, in der Sequenz der zugrunde liegenden Dokumente gedacht, sich in der Nähe zueinander befinden – beides können Hinweise auf einen inneren Zusammenhang der betreffenden Codes, ggf. aber auch auf unzureichende Kategoriendefinitionen sein.

Diese Art grundlegender Quantifizierungen stellt in vielen Fällen neben den klassischen methodenspezifischen Vorgehensweisen einen ersten wichtigen Anlaufpunkt während des Analyseprozesses dar. Sie sind sehr gut geeignet, Ideen bezüglich der Relevanz von Codes oder zu Zusammenhängen zu entwickeln. Sie können Hinweise auf die Güte des Kategoriensystems geben. Sie erlauben es gewissermaßen, die inhaltlich-interpretative Tätigkeit des Codierens aus einem neuen Blickwinkel und im erweiterten Kontext des Gesamtprojekts, aber, wenn gewünscht, auch im Kontext einzelner ausgewählter Dokumente zu betrachten. Die eigentliche Interpretation der Textstellen verliert in dieser distanzierteren Sichtweise zunächst etwas an Relevanz, jedoch nur, um im Anschluss bei der Erschließung der Bedeutungen wieder substanziell in die Analyse einbezogen zu werden.

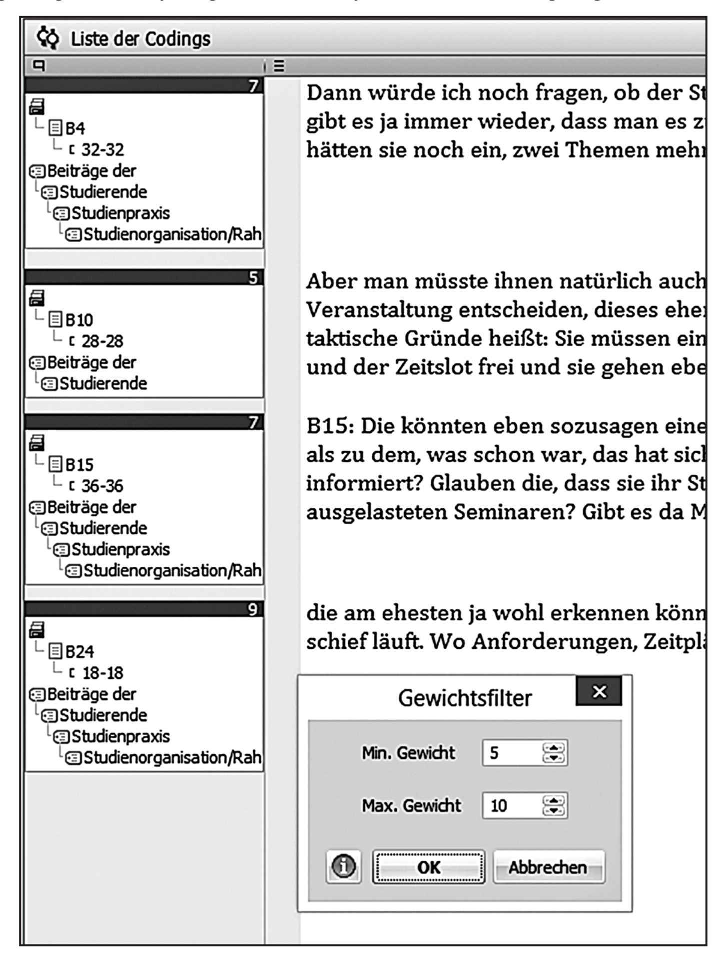
3.2 Gewichtung von Codierungen

MAXQDA bietet die Möglichkeit, jeder Codierung ein eigenes Gewicht zu verleihen. Technisch gesprochen kann so jeder Codierung als Attribut eine Zahl zwischen 0 und 100 zugewiesen werden. Diese Zuweisung stellt einen Interpretationsakt dar und muss demnach für jede Codierung individuell vorgenommen werden. Lediglich das Standardgewicht, also der Wert, der jeder neuen Codierung automatisch zugewiesen wird, lässt sich über den Eintrag *Optionen* im Menü *Projekt* einstellen. Die Voreinstellung 0 wird in der Regel als *nicht gewichtet* verwendet. Das Gewicht einer Textstelle wird verändert, indem der Codierstreifen des betreffenden Codings mit der rechten Maustaste angeklickt wird, woraufhin die Möglichkeit *Gewicht ändern* angeboten wird. Das zugewiesen Gewicht wird unter anderem angezeigt, wenn der Mauszeiger über den Codierstreifen bewegt wird und dort eine Weile ruht, es wird aber auch in den gelben Infoboxen (*Herkunftsangabe*) in der Liste der Codings im oberen blauen Balken ausgegeben.

Anhand der Gewichtungen wird es möglich, die Auswahl anzuzeigender Textstellen differenzierter zu gestalten. Im Menü Analyse werden die Möglichkeiten

geboten, den Gewichtsfilter bearbeiten und den Gewichtsfilter anwenden zu können. Die erste Option gestattet es, eine Wertespanne zu definieren. Die zweite Funktion wendet daraufhin die vorgenommenen Einstellungen auf die Liste der Codings an: Wird der Gewichtsfilter aktiviert, so werden in der Liste der Codings ausschließlich jene Codings der aktivierten Dokumente und Codes gelistet, deren Gewicht innerhalb der definierten Gewichtsspanne liegt. Es handelt sich also um eine effiziente Filtermöglichkeit.

Abb. 4: Beschränkung der anzuzeigenden Codings auf solche mit einem Gewicht der Spanne 5-10. Das tatsächliche Gewicht einer Textstelle wird in der oberen rechten Ecke der zugehörigen Herkunftsangabe (Kasten auf der linken Seite) angezeigt.



Für die praktische Verwendung dieser numerischen Information bieten sich viele Möglichkeiten. Populär ist die Gewichtung beispielsweise als Möglichkeit der Skalierung, etwa auf als Möglichkeit der inhaltlichen Kennzeichnung oder als Umsetzung der *Fuzzy Sets* der QCA (*Qualitative Comparative Analysis*, siehe Ragin 2000). Das erste Beispiel ließe sich etwa dergestalt umsetzen, dass einer Textstelle direkt nach ihrer Codierung ein Attribut in Form des Gewichts zugewiesen wird. Auf diese Weise könnten beispielsweise Textstellen gekennzeichnet werden, die zwar prinzipiell einem Code zuzuordnen sind, ihn aber nicht im Kern treffen oder von minderer inhaltlicher Qualität sind. Ein solches Coding könnte etwa mit einer 1 ausgezeichnet werden. Andersherum könnten auch Textstellen, die einen hervorragenden Beitrag zum jeweiligen Code liefern, ebenso kenntlich gemacht werden, in diesem Beispiel vielleicht mit dem Wert 10. Das zweite Beispiel bezieht sich auf eine eigene analytische Methode, die *QCA* mit *Fuzzy Sets*. Hierbei handelt es sich um ein Verfahren, das zweidimensionale Merkmalsräume definiert. Jedes Dokument oder je nach konkreter Anwendung auch jede einzelne relevante Dokumentpassage wird dann innerhalb dieser Merkmalsräume auf einer prozentualen Skala verortet. In dieser Logik würde das Gewicht diese Verortung verkörpern. So könnte beispielsweise der Grad der Demokratisierung von Ländern erfasst werden: Das Gewicht des Codes *Grad der Demokratisierung* würde für jedes Land abbilden, ob es sich um eine reife, funktionierende und vollständige Demokratie (bspw. Gewicht = 100), ein Land gänzlich ohne demokratische Strukturen (Gewicht = 0) oder vielleicht auch eine Zwischenform mit vielen demokratischen Elementen, jedoch nicht in allen Bereichen der Gesellschaft (bspw. Gewicht = 75) handelt. Die zu vergebenden Gewichtswerte müssen dabei möglichst anhand festgelegter Definitionen reproduzierbar vergeben werden.

Das weitere Verfahren ist prinzipiell identisch: Während des Retrievals, also während des Aufrufens und Anzeigens der codierten Textstellen, kann durch Verwendung des Gewichtsfilters sehr fein bestimmt werden, auf Textstellen welchen Merkmals die Anzeige beschränkt werden soll.

3.3 Dokument- und Codevariablen: Erweiterte Triangulationsmöglichkeiten

Zu jedem Dokument und zu jedem Code können in MAXQDA ergänzende Informationen hinterlegt werden. In der Regel handelt es sich dabei um Daten, die für die Analyse und das Verständnis und die Einordnung der Dokumente zwar relevant sind, selbst aber in MAXQDA nicht oder zumindest

nicht in erster Linie interpretativ bearbeitet werden sollen. Prototypisch für Dokumentvariablen wären etwa soziodemografische Daten sowie weitere relevante Eigenschaften von Interviewpartner/innen, aber auch Ergebnisse einer qualitativen Typenbildung. Codevariablen können beispielsweise im Kontext der *Grounded Theory* (siehe u.a. Corbin/Strauss 2008, Strübing 2008) eingesetzt werden, um die Zugehörigkeit von Konzepten zu Kategorien abzubilden oder, wenn Gruppendiskussionen bearbeitet werden, um die Eigenschaften der beteiligten Sprecher/innen zu erfassen. Da das prinzipielle Vorgehen für die Arbeit mit Dokument- und Codevariablen identisch ist, die Dokumentvariablen aber deutlich häufiger eingesetzt werden und für das Thema der Triangulation auch relevanter sind, fokussiert die nachfolgende Darstellung die Dokumentvariablen.

Neben einigen sog. Systemvariablen, die MAXQDA automatisch zu jedem eingelesenen Dokument erzeugt und die für die interne Datenverwaltung benötigt werden, können Forschende selbst Variablen definieren. Jede Variable steht für jedes Dokument zur Verfügung. Die Darstellung der Daten erfolgt in einer Tabelle gemäß den für solche Datendarstellungen üblichen Konventionen, d.h. dass für jedes Dokument eine Zeile verfügbar ist und für jede existierende Variable eine eigene Spalte erzeugt wird. Um die Liste der Dokumentvariablen aufzurufen, kann der Befehl *Dateneditor (Dokumentvariablen)* im Menü *Variablen* ausgewählt werden. Im sich dann öffnenden Fenster können Werte angesehen, neu eingegeben und verändert werden.

Abb. 5: Ausschnitt des Dateneditors (Dokumentvariablen).

Dokumentvariablen

Alle — 27 Dokumente

Dokument...	Anzahl Codings	Anzahl Memos	Informat...	Statusgruppe	Methode Le...	Methode Stu...	Art des Einbe...	M
B1	47	2	Ja	Lehrende	Standardisiert	Standardisiert	beteiligt	St
B2	48	1	Ja	Studierende	Mixed	Mixed	beteiligt	St
B3	52	6	Ja	Lehrende	Mixed	Offen	beteiligt	
B4	44	1	Ja	Lehrende	Standardisiert	Standardisiert	beteiligt	
B5	51	1		Studierende	Mixed	Standardisiert	beteiligt	M
B6	37	1	Ja	Lehrende	Offen	Standardisiert	beteiligt	
B7	44	1	Ja	Studierende	Mixed	Mixed	beteiligt	M
B8	42	1	Ja	Studierende	Standardisiert	Standardisiert	beteiligt	St
B9	38	3	Nein	Studierende	Offen	Standardisiert	beteiligt	O
B10	45	3	Ja	Lehrende	Mixed	Mixed	beteiligt	M
B11	37	5	Nein	Studierende		Standardisiert	beteiligt	O
B12	47	3	Ja	Lehrende	Mixed	Mixed	beteiligt	M
B13	46	7	Nein	Studierende	Offen	Offen	betroffen	O
B14	41	5		Studierende	Offen	Mixed	betroffen	
B15	37	1		Studierende	Mixed	Standardisiert	betroffen	St
B16	44	5		Studierende	Mixed	Standardisiert	betroffen	
B17	40	4		Studierende	Mixed	Mixed	betroffen	

Um neue Variablen zu definieren, wird über den Eintrag *Liste der Dokumentvariablen*, der sich ebenfalls im Menü *Variablen* befindet, der entsprechende Dialog aufgerufen[2]. Wird das Icon zur Erzeugung einer *Neuen Variable* angeklickt, muss als nächstes definiert werden, welchen Namen die neue Variable erhalten soll und um welchen Datentyp es sich handelt. Zudem ist es möglich, eine Zeichenkette zu hinterlegen, anhand derer fehlende Werte identifiziert werden. Der Name der Variable kann frei gewählt werden. Der Typ einer Variablen definiert, welche Art von Daten die neue Variable aufnehmen kann. Sollen neben Ziffern auch weitere Zeichen wie Buchstaben oder Sonderzeichen hinterlegt werden, so muss als Typ *String* gewählt werden, da außer diesen Stringvariablen keine andere Variable Buchstaben entgegen nimmt. Daneben sind noch die Variablentypen *Ganzzahl* zur Speicherung von Ziffern (etwa Geburtsjahr), *Fließkommazahl* zum Hinterlegen von Werten mit Nachkommastellen (beispielsweise die Abiturdurchschnittsnote), *Datum/Uhrzeit* sowie *Boolean* verfügbar. Der Variablentyp *Boolean* ist zweiwertig, d.h. er ermöglicht die Hinterlegung von lediglich zwei unterschiedlichen Werten: *Vorhanden* bzw. e*rfüllt* und *nicht vorhanden* bzw. *nicht erfüllt*. Während für alle anderen Variablentypen ein Eingabefeld bereitgestellt wird, in das die betreffenden Werte eingetragen werden können, wird für *Boolean*-Variablen eine Checkbox dargestellt, die abgehakt werden kann. *Vorhanden* bzw. *erfüllt* wird dann durch ein gesetztes Häkchen dargestellt. Diese Variablenart eignet sich demnach für die Erfassung von Merkmalen, die entweder zutreffen oder eben nicht, also beispielsweise das Funktionsmerkmal *Schulleiter/in* oder die Zuordnung einer interviewten Person zum während der Analyse gebildeten Typus *Verweigerer*.

Nicht zufällig ähneln diese Beschreibung und die verwendeten Begriffe Verfahren aus der Welt der statistischen Datenanalyse: Die Variablen stellen einen zentralen Verbindungspunkt zwischen qualitativen und quantitativen Daten in MAXQDA dar. Dabei geht es nicht nur darum, die Informationen einfach in einer Datei zusammenzuführen. Vielmehr wird der Ansatz verfolgt, das qualitative und das quantitative Methodenspektrum möglichst eng und unkompliziert, gleichzeitig aber so flexibel wie möglich zusammenzuführen. Hierzu gehört unter anderem, dass es nicht erforderlich ist, bereits in Tabellenform vorliegende Daten – z.B. aus SPSS, Stata oder Excel – in MAXQDA erneut einzugeben, sondern es ist möglich, sie mithilfe der Funktionen *Daten exportieren (Dokumentvariablen)* und *Daten importieren (Dokumentvariablen)* im Menü

---

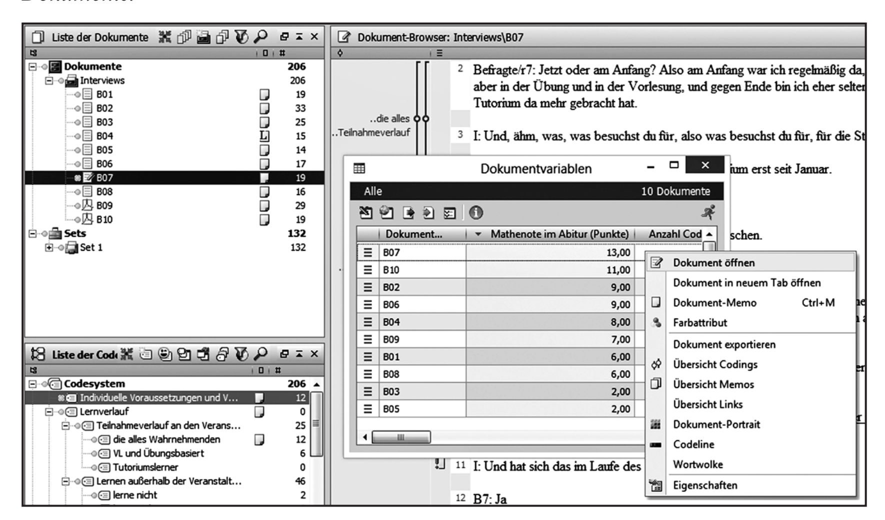
2 Für viele Funktionen bietet MAXQDA mehr als eine Möglich wird stets nur eine Zugriffsmöglichkeit beschrieben.

Variablen unkompliziert nach MAXQDA zu transferieren. Das Spektrum der Verwendungsmöglichkeiten ist sehr breit gestreut, viele der komplexeren Analysetechniken, die in MAXQDA zur Verfügung stehen, basieren entweder auf Variablen oder sind in der Lage, die in der Variablenliste hinterlegten Werte in die Auswertung einbeziehen. Nachfolgend werden die wichtigsten Einsatzmöglichkeiten der Variablen vorgestellt, die sich im Kontext vieler Forschungsprojekte als fruchtbar erweisen können.

Variablenwerte als ergänzende Information und als Grundlage manueller Auswahl

Die einfachste Möglichkeit, die Variablenwerte für die Analyse zu nutzen, ist, sie als ergänzende Kontextinformation zu betrachten und bei Bedarf aufzurufen. Dazu kann aus dem Kontextmenü eines Dokumentes, das mit einem Rechtsklick auf das betreffende Dokument aufgerufen wird, direkt der Eintrag *Übersicht Variablen* ausgewählt werden. Im sich öffnenden Fester werden alle zu diesem spezifischen Dokument hinterlegten Variablenwerte angezeigt. Dieser Weg führt sozusagen von den qualitativen Daten zu den quantitativen Variablenwerten. Doch auch in umgekehrter Logik, also in Richtung von den Variablen zu den qualitativen Daten, lassen sich die Variablen sehr einfach einsetzen. In diesem Fall würden die Variablen als Kriterium zur manuellen Auswahl von Fällen mit interessanten und relevanten Merkmalen herangezogen. Ausgangspunkt ist dann der *Dateneditor (Dokumentvariablen)*. Klickt man mit linker Maustaste auf einen der Spaltenköpfe, die den Namen der Variablen enthalten, so wird die Liste nach den Werten dieser Spalte aufsteigend sortiert, ein erneuter Klick kehrt die Sortierreihenfolge um. So könnten die Dokumente etwa nach dem Alter der datengebenden Personen sortiert werden. Auf diese Weise wird das Spektrum der hinterlegten Werte sehr schnell sichtbar. Um ein Dokument nun zu betrachten oder zu analysieren, klickt man mit der rechten Maustaste auf die entsprechende Zeile und wählt den Eintrag *Dokument öffnen*. Auf diese Weise lässt sich beispielsweise einfach und schnell untersuchen, ob die Vermutung, dass bestimmte in Variablen erfasste Merkmale mit inhaltlichen Äußerungen der Befragten in Verbindung steht, weiter verfolgt werden sollte oder nicht. Letztlich handelt es sich um nichts anderes als eine einfache Form von Gruppenvergleichen bzw. einer tendenziell eher explorativen Suche nach Mustern.

Abb. 6: Aufruf eines Dokumentes über das Fenster Dateneditor (Dokumentvariablen). Die Einträge des Dateneditors wurden durch Klick auf die Spaltenüberschrift nach der Variable Mathenote im Abitur (Punkte) absteigend sortiert, so dass die Personen mit der höchsten Punktzahl nach oben und jene mit der niedrigsten Punktzahl unten sortiert wurden. Ein Rechtsklick auf eine Zeile ermöglicht den direkten Aufruf eines interessierenden Dokuments.



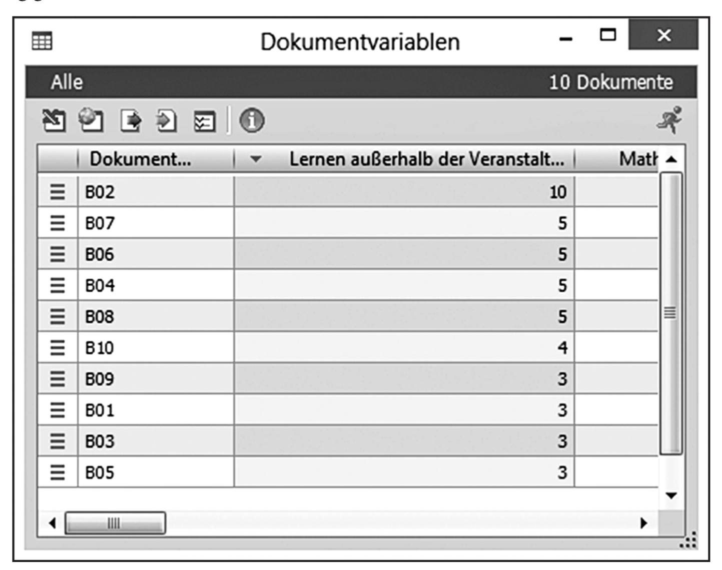
Aktivierung via Dokumentvariablen
In die gleiche Richtung weisend, hinsichtlich der sich bietenden Möglichkeiten jedoch deutlich weitreichender, ist die Funktion *Aktivierung via Dokumentvariablen* (in MAXQDA vor Version 11 lautet der Name dieser Funktion *Logische Aktivierung*). Dabei handelt es sich um die Möglichkeit, automatisiert aus der Liste der Dokumente all jene zu aktivieren, die den vorgegebenen Kriterien entsprechen. Als Kriterien, die zur Auswahl herangezogen werden können, dienen wiederum die hinterlegten Variablenwerte. Damit ermöglicht diese Funktion u.a. einen Rückgriff auf standardisierte, quantitative Informationen als Auswahlkriterium.

Die *Aktivierung via Dokumentvariablen* kann über das Menü *Mixed Methods* aufgerufen werden. Im sich öffnenden Dialog werden eine oder mehrere Bedingungen definiert, auf die hin der Datenkorpus überprüft werden soll. Entspricht ein Dokument den vorgegebenen Bedingungen, so wird es aktiviert, andernfalls deaktiviert. Dabei kann auf sämtliche im Projekt erzeugten Variablen zurückgegriffen werden. Zudem können mehrere Variablenbedingungen mit den Operanden *und* sowie *oder* miteinander verknüpft werden, so dass auch sehr

komplexe Bedingungen abgebildet werden können. Unter der Voraussetzung, dass die entsprechenden Informationen als Variablen hinterlegt wurden, ist es mithilfe dieser Funktion ein Leichtes, die Analyse beispielsweise auf Dokumente der Angehörigen einer bestimmten Partei in einer bestimmten Funktion in einem bestimmten politischen Gremium zu beschränken und im nächsten Schritt eine ganz andere Gruppe zu wählen. Vor allem in Verbindung mit der Möglichkeit, die aktivierten Dokumente in sog. *Sets*, in frei zu erzeugenden Gruppen von Dokumenten zu organisieren, ermöglicht diese Funktion sehr elegant, Dokumente gezielt und flexibel gruppenspezifisch für die Analyse auszuwählen. Dabei ist diese Funktion nicht nur bei großen Projekten mit umfangreichen Datenkorpora hilfreich, in denen eine manuelle Auswahl aufgrund der schieren Masse der Dokumente unpraktikabel wäre. Durch die Möglichkeit der Verknüpfung mehrerer Bedingungen bietet sich ihr Einsatz auch in Projekten mit nicht zu vielen zu analysierenden Dokumenten an und erspart langwierige manuelle Wertevergleiche.

Umwandlung von Codes in Dokumentvariablen
Neben der Möglichkeit, selbst neue Variablen zu erzeugen und Daten einzugeben, erlaubt MAXQDA auch, Codes automatisiert in Dokumentvariablen zu transformieren. Dazu wird der Eintrag *In Dok.-Variable transformieren* aus dem Menü aufgerufen, das erscheint, wenn mit der rechten Maustaste auf einen Code geklickt wurde. Im Anschluss öffnet sich der *Dateneditor (Dokumentvariablen)* und zeigt die neu erzeugte Spalte ganz am rechten Rand der Tabelle an. Als Überschrift, also als Variablennamen, trägt die Spalte den Namen des Codes. Darunter trägt MAXQDA für jedes Dokument als Zahl ein, wie häufig der betreffende Code im jeweiligen Dokument vergeben wurde. Damit steht dieser Wert direkt als mögliches Kriterium zur Dokumentenauswahl zur Verfügung, da beispielsweise die *Aktivierung via Dokumentvariablen* auf Basis dieser Werte umgesetzt werden können. Beispielsweise ließen sich nun automatisiert alle Dokumente mithilfe der *Aktivierung via Dokumentvariablen* wählen, die besonders viele oder eben besonders wenige Codings zum betreffenden Code beinhalten. Diese Funktion ermöglicht also die Auswahl von Dokumenten aufgrund von quantifizierten Codiermerkmalen.

Abb. 7: Ergebnis der Umwandlung des Codes in eine Dokumentvariable: Für jedes Dokument wird in der Variablenliste vermerkt, wie häufig der umgewandelte Code im betreffenden Dokument vergeben, d.h. wie viele Dokumentstellen mit ihm codiert wurden. Die Werte einer solchen Variable können für alle hier beschriebenen Operationen vollwertig genutzt werden.

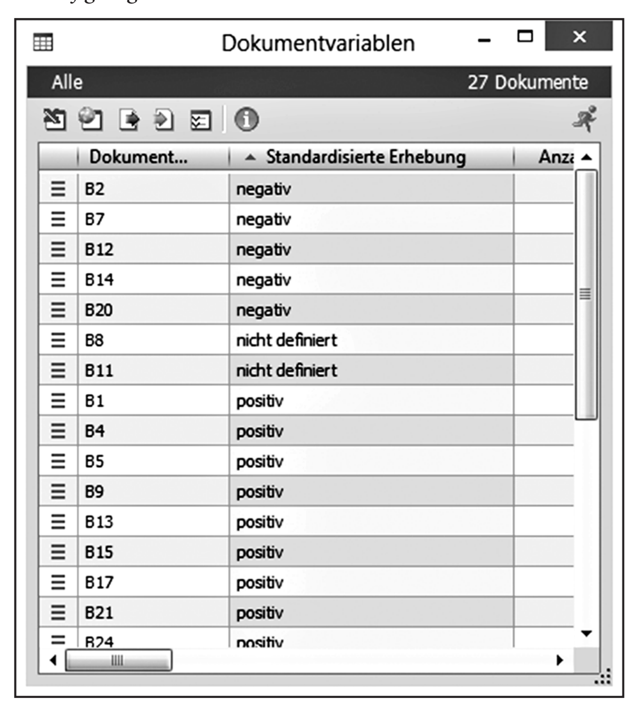


Umwandlung von Codes in kategoriale Dokumentvariablen

Kategoriale Variablen bieten die Möglichkeit, Bewertungen, die in Form von Subcodes vorgenommen wurden, als Variablenausprägung abzubilden und so für die weitere Analyse mit den oben beschriebenen Methoden zugänglich zu machen. Voraussetzung ist ein Code, dem nominal oder auf einer höheren Stufe skalierte Subcodes zugeordnet wurden. So könnten beispielsweise dem Code *Bewertung von Veranstaltung und Inhalt* in einer qualitativen Evaluationsstudie die Subcodes *Positiv*, *Neutral* sowie *Negativ* zugeordnet werden. Während des Codierens würden nun relevante Textstellen nicht dem Obercode zugewiesen, sondern die codierende Person würde direkt beurteilen, welche Bewertungstendenz eine Textstelle ausdrückt und sie dann direkt dem entsprechenden Subcode zuweisen. Um diese Codierungen als Variablenwerte zugänglich zu machen, wird mit einem Rechtsklick auf den Hauptcode *Bewertung von Veranstaltung und Inhalt* das Kontextmenü aufgerufen, aus dem dann der Eintrag *In kategoriale Dok.-Variable transformieren* aufgerufen wird. Es öffnet sich der *Dateneditor (Dokumentvariablen)* und zeigt am rechten Ende die neu erzeugte

Spalte, die wiederum den Namen des Codes als Bezeichnung trägt. Als Wert wird die Bezeichnung des Subcodes von *Bewertung von Veranstaltung und Inhalt* ausgegeben, der in dem betreffenden Dokument am häufigsten vergeben wurde, also entweder *Positiv*, *Neutral* oder *Negativ*. Wurden mindestens zwei Subcodes gleich häufig vergeben, so wird als Wert *nicht definiert* eingetragen. In diesen Fällen müssen die Forscher/innen die entsprechend codierten Segmente inspizieren und eine Entscheidung über den adäquaten Variablenwert fällen. Diese Funktion klassifiziert demnach qualitative Bewertungen aufgrund ihrer Häufigkeiten. Auf dieser Basis wäre es nun ohne weiteres möglich, die Analyse mithilfe der Funktion *Aktivierung via Dokumentvariablen* auf jene Dokumente zu begrenzen, deren Datengeber/innen vornehmlich negative Bewertungen abgegebenen haben, um sie im nächsten Schritt mit jenen zu vergleichen, die sich mehrheitlich positiv geäußert haben.

Abb. 8: Ergebnis der Umwandlung eines Codes in eine kategoriale Variable: In die automatisch während der Umwandlung erzeugte Variable mit dem Namen des Codes (Standardisierte Erhebung) wird für jedes Dokument eingetragen, welcher der Subcodes in dem betreffenden Dokument am häufigsten vergeben wurden. Nicht definiert bedeutet, dass mindestens zwei Subcodes gleich häufig vergeben wurden.

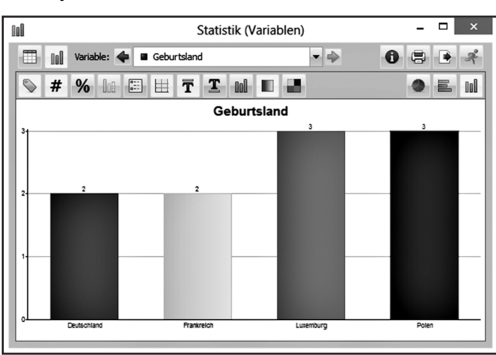


Dokument...	Standardisierte Erhebung	Anz
B2	negativ	
B7	negativ	
B12	negativ	
B14	negativ	
B20	negativ	
B8	nicht definiert	
B11	nicht definiert	
B1	positiv	
B4	positiv	
B5	positiv	
B9	positiv	
B13	positiv	
B15	positiv	
B17	positiv	
B21	positiv	
B24	positiv	

Statistiken
Eine in MAXQDA 11 neu hinzugekommene Möglichkeit, mit Quantifizierungen zu arbeiten, stellen *Statistiken* dar. Diese Funktion ermöglicht es, die Daten, die in der Variablenliste hinterlegt wurden, direkt in MAXQDA mithilfe von Methoden deskriptiver Statistik zu bearbeiten, um für einfache Betrachtungen den Weg zur Statistik-Software vermeiden zu können. Zu Verfügung stehen absolute und relative Häufigkeiten, Mittelwerte und die Standardabweichung. Zudem können verschiedene Diagramme erzeugt und in ihrem Erscheinungsbild verändert werden. Neben der Möglichkeit, standardisierte Informationen schnell und direkt innerhalb der QDA-Software auszuwerten, bietet sich diese Verfahrensweise im Sinne der Triangulation auch an, um qualitative Informationen auszuwerten, die während der Analyse erzeugt und als Variable hinterlegt wurden, ohne ein weiteres Analyseprogramm bemühen zu müssen. Auch hier wäre ein klassisches Beispiel die Zuordnung von Dokumenten zu im Rahmen einer Typenbildung identifizierten und als Variablenausprägungen festgehaltenen Typen. Um solche Statistiken für Variablen zu erzeugen, kann der Menüpunkt *Statistik für Dokumentvariablen* im Menü *Variablen* aufgerufen werden.

Neben der Möglichkeit, Variablenwerte auf diese Weise auszuwerten, ist es auch möglich, eine differenzierte quantifizierte Übersicht über Subcodes zu erzeugen. Dazu wird über das Menü *Mixed Methods* oder das Menü *Codes* der Eintrag *Statistik für Subcodes* ausgewählt. Im Anschluss muss angegeben werden, für welche Hauptcodes Statistiken erzeugt werden sollen. Anschließend werden die Subcodes des oder der gewählten Codes in Tabellenform aufgelistet. Zu jedem Subcode werden die absolute und die relative Häufigkeit seiner Anwendung im Projekt oder, je nach Vorauswahl, in Teilen des Projektes dargestellt. Auch hier können, wie bei den anderen Statistikfunktionen, auf Wunsch Diagramme der Verteilungen erzeugt werden.

Abb. 9: Säulendiagramm für die Variable Geburtsland.



3.4 Joint Displays

Der Begriff *Joint Displays* bezeichnet, zurückgehend auf Creswell/Plano-Clark (2011: 226), die gemeinsame Darstellung von qualitativen und quantitativen Daten in einer Tabelle oder Grafik, um sie direkt vergleichen zu können. Sie sind somit ein zentrales Element der Zusammenführung und Auswertung von Informationen unterschiedlicher Art und Herkunft. MAXQDA 11 ermöglicht die Erzeugung von vier unterschiedlichen Joint Displays: *Segmentmatrizen*, *Kreuztabellen*, *Typologietabellen* sowie *Konfigurationstabellen*.

Segmentmatrizen

Als *Segmentmatrix* (über den entsprechenden Eintrag im Menü *Mixed Methods* aufzurufen) wird in MAXQDA eine Zusammenstellung von Codings bezeichnet, die nach in Form von Dokumentvariablen hinterlegten Eigenschaften sortiert nebeneinander gestellt werden. Beispielsweise ließen sich die Aussagen, die im Rahmen einer Lehrevaluation zur Verbesserung von Veranstaltung und Inhalt der Veranstaltung geäußert wurden, danach aufteilen, ob die sich äußernden Personen die Abschlussklausur eben dieser Veranstaltung bestanden haben oder nicht – vorausgesetzt, die Information bzgl. des Bestehens wurde als Dokumentvariable erfasst. Der Variable *Bestanden* käme somit eine Filterfunktion zu. Im Ergebnis, als eigentliche Segmentmatrix, wird den Forschenden eine Tabelle präsentiert. In der ersten Spalte findet sich eine Liste der ausgewählten Codes, wobei jeder Code und jeder Subcode eine eigene Zeile erhalten. In den nachfolgenden Spalten werden die Textstellen nach gewählten Filtermerkmalen sortiert dargestellt. So würde die zweite Spalte in unserem Beispiel die Aussagen all jener Personen enthalten, welche die Klausur bestanden haben, während die dritte Spalte alle Aussagen von Personen bündelt, welche die Klausur nicht bestanden haben. Die Codings lassen sich also nach Variablenmerkmalen gruppiert darstellen. Diese gemeinsame Darstellung ermöglicht beispielsweise einen direkten Vergleich der Aussagen zweier als möglicherweise relevant eingeschätzten Gruppen von Merkmalsträger/innen, indem qualitative und quantitative Informationen in einer Tabelle zusammengeführt werden.

Abb. 10: Ausschnitt einer Segmentmatrix. Dargestellt werden Codings zum Code Verbesserungsvorschläge, *wobei die Aussagen der Personen, welche die Klausur bestanden haben, in der zweiten Spalte und die derjenigen, die die Klausur nicht bestanden haben, in der dritten Spalte dargestellt werden.*

|	Klausur bestanden = 1	Klausur bestanden <> 1
Verbesserungs vorschläge	Ich finde, ähm, vielleicht könnte man, kann man, also wenn es möglich ist, vielleicht kann man diese Übung auf zwei Gruppen aufteilen und dann könnte man diese, die Fragen, die dann entstehen, eher noch durchgehen, als so in diesem großen Raum, also in der großen Gruppe ist das wahrscheinlich schwierig Interviews\B03, 66 B4: Also es ist jetzt für mich jetzt nen bisschen bisschen schwierig, weil ääh, ich bin jetzt ein Semester hier und ich weiß nicht also es ist ein bisschen, also ich weiß jetzt nicht ob ich da was sagen sollte was oder kann überhaupt also, wie gesagt, wie fands gut die Sachen die die mir nicht gefallen haben das sind ja so Sachen die mir die jetzt nicht so allgemein verändert werden werden müssen, wie zum Beispiel das mit dem Tutorium oder was, also da ist jetzt oder Vorlesung das ist ja ist ja so, also ich hab jetzt nicht so wirklich nen Verbesserungsvorschlag. Interviews\B04, 34 I: Damit eben beschäftigt sein und was	B6: (...). An sich haben mir die Veranstaltungen eigentlich gut gefallen nur der Lärmpegel war manchmal doch etwas zuu laut I: ja B6: dass da dann irgendwie das nächste ma ein bisschen I: Ja B6: entgegengewirkt I: Ja B6: wird. Interviews\B06, 18-24 B6: Verbesserungswünsche eigentlich nur, dass man irgendwie was dann gegen den Lärm macht Interviews\B06, 32

Kreuztabellen

Prinzipiell folgen die *Kreuztabellen* (aufzurufen über das Menü *Mixed Methods*) der gleichen Logik wie Segmentmatrizen, d.h. hier werden Codings auf der Grundlage von in Dokumentvariablen vorliegenden Werten gruppiert und dargestellt. Allerdings werden in Kreuztabellen nicht die tatsächlichen Dokumentstellen ausgegeben, sondern die Darstellung erfolgt in aggregierter Form: Angezeigt werden die absoluten oder wahlweise relativen Häufigkeiten der Vergabe eines bestimmten Codes in allen Dokumenten, aufgeteilt nach gewählten Variablenmerkmalen. Würde das bei den Segmentmatrizen vorgestellte Beispiel als Kreuztabelle dargestellt, so würde in der ersten Spalte der erzeugten Tabelle in jeder Zeile ein Code aufgeführt. Die zweite und dritte Spalte wären ebenfalls mit den Merkmalsausprägungen *Bestanden* bzw. *Nicht bestanden* überschrieben. In diesen zweiten und dritten Spalten würde nun als Zahlenwert ausgegeben, wie viele Stellen in Dokumenten, die dem jeweiligen Kriterium entsprechen, dem entsprechenden Code

zugewiesen wurden. Kreuztabellen sind gleichsam eine quantitative Form der Segmentmatrizen und erlauben somit ebenfalls einen sehr effizienten Vergleich von Gruppen, nun allerdings auf quantifizierter, nicht auf qualitativer Ebene.

Abb. 11: Kreuztabelle zum Code Verbesserungsvorschläge, *Gruppierung der Codehäufigkeiten nach der* Variable Klausur bestanden.

Kreuztabellen

Codesystem	Klausur bestanden = 1	Klausur bestanden <> 1	SUM
Verbesserungsvorschläge	4	2	6
Irritation bei Folien	3		3
kein Lärm zulassen	3		3
Reader	1	1	2
SPSS für alle		2	2
mehr Zeit bei Übung	1		1
weniger Stoff	1		1
mehr mit Praxis verknüpfen	1		1
aufpeppen mit Medien	1		1
mehr Tutorien		1	1
Σ SUM	15	6	21
# N (Dokumente)	5	5	

Typologietabellen

Während Segmentmatrizen und Kreuztabellen auf der Integration von Variablen und Codes basieren, ermöglichen *Typologietabellen* (ebenfalls über das Menü *Mixed Methods* aufzurufen) die in einer Tabelle vereinte Gruppierung von Variablenwerten und anderen Variablenwerten. Zur Verdeutlichung soll angenommen werden, dass neben der Tatsache, ob eine befragte Person die Klausur bestanden hat oder nicht, auch als Variable erfasst worden wäre, mit welchen Strategien sie außerhalb des Besuchs der Veranstaltungen gelernt hat. Dabei ist zweitrangig, in welcher Form diese Lernstrategien faktisch in der Variablentabelle festgehalten werden, da die Funktion der Typologietabellen sowohl auf numerische als auch auf Stringvariablen anwendbar ist. Für die weitere Analyse wäre es nun hilfreich zu wissen, ob eine Häufung zwischen bestimmten Lernstrategien und dem Erfolg des Absolvierens der Klausur festgestellt werden kann. Genau dieser Frage lässt sich mithilfe der Typologietabelle nachspüren, da sie ermöglicht, summarisch darzustellen, mit welchem Erfolg die Nutzer/innen welcher Lernstrategie die Klausur abgeschlossen haben. Dazu werden nach dem Aufruf dieser Funktion zunächst aus einer Liste der existierenden Variablenwerte die hier näher zu betrachtenden ausgewählt, in diesem Fall also die

erfassten Lernstrategien. Diese Werte stellen in der späteren Tabelle die Zeilen dar, jede gewählte Lernstrategie wird in der resultierenden Tabelle in einer eigenen Zeile ausgegeben. Im nächsten Schritt ist auszuwählen, nach welchen Werten die Spaltenaufteilung erfolgen soll. Hier wäre demnach die Variable *Klausur* mit ihren Werten *Bestanden* und *Nicht bestanden* auszuwählen. Als Ergebnis wird eine Tabelle angezeigt, in der man sofort ersehen kann, welcher Anteil der Nutzer/innen eines Lernweges die Klausur bestanden haben oder nicht. Dabei werden je nach Datentyp der Variable unterschiedliche Werte mit ausgegeben, nämlich wahlweise die absoluten und relativen Häufigkeiten oder, bei numerischen Variablen, Mittelwert und Standardabweichung. Somit stellt die Typologietabelle ein Werkzeug dar, mit dessen Hilfe sich Merkmalshäufungen im auszuwertenden Material darstellen lassen. Sie ist eine wichtige Hilfe bei der Suche nach Mustern sowie bei der Suche nach unterscheidbaren Typen.

Abb. 12: Typologietabelle für die Variable Lernstrategien, aufgeteilt nach der Variable Klausur bestanden.

Typologietabelle

	Klausur bestanden = 1 (N=5)	Klausur bestanden <> 1 (N=5)
Lernstrategie: Nicht-Lerner_in, Anzahl (%)	0	2 (40,0)
Lernstrategie: Selbstlerner_in, Anzahl (%)	3 (60,0)	1 (20,0)
Lernstrategie: Veranstaltungslerner_in, Anzahl (%)	2 (40,0)	2 (40,0)

Konfigurationstabellen

Die *Konfigurationstabelle* (Menü *Mixed Methods*) erlaubt die Beantwortung der Frage, wie häufig ausgewählte Codes gemeinsam in einem Dokument vergeben wurden. Dabei meint *gemeinsam* in diesem Falle nicht *gleichzeitig* im Sinne einer Überschneidung oder *nah* im Sinne der Nähenanalyse des Code-Relations-Browsers, sondern es geht um die Frage, ob die zu untersuchenden Codes irgendwo im gleichen Dokument vergeben wurden – also beispielsweise darum, welche Personen sich konkret zu welchen Themen geäußert haben. Nach dem Aufruf der Funktion müssen zunächst die Codes ausgewählt werden, die untersucht werden sollen. Daran anschließend werden zwei tabellarische Übersichten präsentiert, die *Konfigurationstabelle* und die *Detailansicht*. In der *Konfigurationstabelle* wird dargestellt, welche der theoretisch möglichen Codekombination im vorliegenden Material tatsächlich vorzufinden sind. Zusätzlich wird ausgegeben, mit welchen absoluten und relativen Häufigkeiten jede der existierenden Kombinationen auftritt. In der *Detailansicht* wird auf Dokumentebene dargestellt, welcher dieser Kombinationen jedes der im Projekt

enthaltenen Dokumente zugehörig ist. Als Beispiel ist davon auszugehen, dass die im vorigen Absatz bereits erwähnten Lernstrategien im Dokument selbst codiert worden wären, dass also ein Code *Lernstrategien* mit den Subcodes *Lesen, Übungsaufgaben bearbeiten, Selbstlernen* und *Lerngruppe* existiert. Werden die vier interessierenden Subcodes in Form einer Konfigurationstabelle zusammengeführt, so wird ersichtlich, welche Lernstrategien jede der befragten Personen verfolgt hat und wie sich das Verhältnis der verschiedenen möglichen Kombinationen von Lernstrategien darstellt. Somit eignen sich Konfigurationstabellen hervorragend, um nach individuellen Mustern von Inhalten und Eigenschaften zu suchen, die in Form von Codes in den Dokumenten erfasst wurden.

Abb. 13a: Konfigurationstabelle: Es wird dargestellt, dass sieben der 16 theoretisch möglichen Codekombinationen (siehe obere rechte Ecke des Fensters) im Material vorkommen, um welche es sich handelt und wie häufig sie im gesamten codierten Material auftreten. Drei Personen lösen ausschließlich Übungsaufgaben, zwei lesen im Selbststudium etc.

Konfigurationstabelle

7 (16) Kombinationen

	Lesen	Übungsaufg...	Selbstlernen	Lerngruppe	Häufigkeit	Prozent
•		■			3	30,00
•	■		■		2	20,00
•	■				1	10,00
•	■	■		■	1	10,00
•	■	■	■	■	1	10,00
•			■		1	10,00
•		■		■	1	10,00
Σ					10	100,00

Abb. 13b: Detailansicht der Konfigurationstabelle: Es wird dargestellt, welche Merkmalskombination (hier: verfolgte Lernstrategien) in welchem Dokument vorzufinden ist.

Konfigurationstabelle - Detailansicht

10 Dokumente

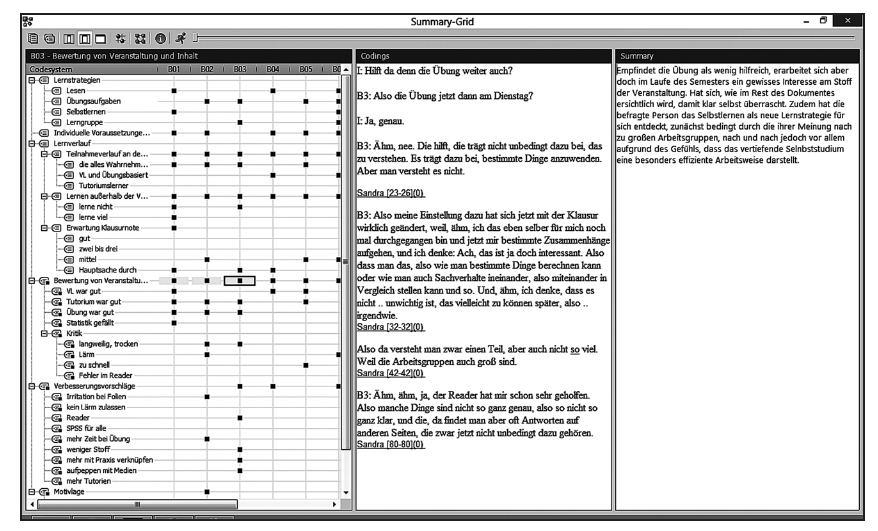
	Dokument...	Dokument	Lesen	Übungsaufg...	Selbstlernen	Lerngruppe	Summe
•	Interviews	B01	■		■		2
•	Interviews	B02		■			1
•	Interviews	B03		■		■	2
•	Interviews	B04	■				1
•	Interviews	B05		■			1
•	Interviews	B06	■	■	■	■	4
•	Interviews	B07	■		■		2
•	Interviews	B08	■	■		■	3
•	Interviews	B09			■		1
•	Interviews	B10		■			1

3.5 *Summary Grids*: Interpretationen und Variablen gleichzeitig darstellen

Die Funktion der *Summary Grids* ist ebenfalls eine Neuerung in MAXQDA 11. Sie ermöglicht in einem zweischrittigen Vorgehen, für jedes einzelne Dokument zunächst eigene interpretative und/oder zusammenfassende Texte zu allen einem Code zugewiesenen Passagen zu verfassen, um diese selbst verfassten Deutungen und Synopsen im nächsten Schritt kumuliert darzustellen und, wenn gewünscht, mit in Variablen hinterlegten Informationen zusammenzuführen. Um die Zusammenfassungen zu erstellen, muss aus dem Menü *Summaries* der Eintrag *Summaries erstellen und editieren* ausgewählt werden. Im erscheinenden Fenster ist es nun möglich, durch einen Klick im linken Fensterdrittel (Tabelle mit Codes in den Zeilen und Dokumenten in den Spalten) alle einer bestimmten Kategorie zugewiesenen Passagen eines Textes im mittleren Fensterteil (*Codings*) anzeigen zu lassen. Im rechten Fensterteil (*Summary*) kann nun ein freier Text zu den angezeigten Codings verfasst werden. Der hier eingegebene Text wird von MAXQDA intern mit den angezeigten Textstellen verbunden. Diese Funktion ist demnach ideal dazu geeignet, code- und gleichzeitig dokumentenspezifische Zusammenfassungen und Deutungen zu erstellen. Um diese selbst produzierten Texte flexibel und in unterschiedlichsten Zusammenstellungen wieder auszugeben, werden im zweiten Schritt *Grid-Tabellen* erstellt. Diese Funktion wird ebenfalls über das Menü *Summaries* aufgerufen. Eine Grid-Tabelle stellt die gewählten Summaries gemeinsam mit frei wählbaren Werten aus der Variablenliste in einer gemeinsamen Tabelle dar. Beim Erzeugen einer neuen Grid-Tabelle muss zunächst ausgewählt werden, ob Variablenwerte mit in die Darstellung einbezogen werden sollen und wenn ja, welche. Danach ist auszuwählen, welche Interpretations- bzw. Zusammenfassungstexte in die Tabelle aufgenommen werden sollen. Die Auswahl erfolgt dabei anhand der Codes des Codesystems: Wird ein Code ausgewählt, so werden alle Summaries, die im vorherigen Schritt zu den Codings der Dokumente erstellt wurden, in die Tabelle eingefügt. Wäre also zu jedem Dokument der oben bereits erwähnten fiktiven Lehrevaluation eine Zusammenfassung zu den Textstellen erstellt worden, die dem Code *Bewertung von Veranstaltung und Inhalt* zugeordnet wurden, so wäre es nun möglich, all diese Zusammenfassungen gemeinsam auszugeben. Dazu müsste bei der Erstellung des Summary-Grids lediglich der Code *Bewertung von Veranstaltung und Inhalt* ausgewählt werden. Um erläuternde Variablenwerte mit in der Tabelle auszugeben, könnten zudem etwa die Variablen *Abiturnote* oder *Fachsemester* ausgewählt werden, die dann in der Ergebnistabelle neben den Zusammenfassungen ausgegeben würden.



Abb. 14: Erzeugen von Summaries.



3.6 Die Verwendung von Werkzeugen der quantitativen Inhaltsanalyse: Worthäufigkeiten und MAXDictio

Neben den latenten Sinngehalten, die sich mithilfe interpretativ-qualitativer Verfahren erschließen lassen, sind mitunter auch quantitative Merkmale der zu analysierenden Dokumente von Interesse. So können beispielsweise die Begriffe, mit denen befragte Personen bestimmte Gegenstände bezeichnen, einen Rückschluss auf ihre Sichtweise ermöglichen. So macht es etwa einen Unterschied, ob Personen im Kontext sozialer Arbeit von *Klienten* oder *Kunden* sprechen oder ob vornehmlich das Wort *Bedürftige* verwendet wird. MAXQDA ermöglicht, mithilfe der Funktion *Worthäufigkeiten* einen schnellen und genauen Überblick über den Wortschatz eines oder mehrerer Dokumente zu gewinnen und Wortschatzanalysen durchzuführen. Die Funktion wird über das Menü *MAX-Dictio* aufgerufen. Bevor die eigentliche Liste präsentiert wird, können diverse Einstellungen vorgenommen werden, die sich auf das Zählergebnis auswirken. So kann die Analyse etwa auf einige aktivierte Dokumente oder auch auf die *Liste der Codings* beschränkt werden. Außerdem ist es möglich, eine *Stoppliste* auszuwählen, die Wörter enthält, die bei der Analyse ignoriert werden. So kann es beispielsweise oftmals zweckmäßig sein, Konjunktionen und Artikel sowie nicht sinntragende Wörter von der Auszählung auszuschließen. Die Liste der Worthäufigkeiten selbst enthält schließlich alle in den gewählten Dokumenten

vorkommenden Worte, die nicht über die Stoppliste von der Auszählung ausgeschlossen wurden. Neben dem Wort selbst werden seine Länge sowie seine absolute und relative Häufigkeit innerhalb der Gruppe der für die Analyse gewählten Dokumente ausgegeben. Wieder ist es möglich, durch einen Klick auf den Spaltenkopf nach jeder dieser Spalten zu sortieren, so dass eine alphabetische Sortierung der Fundworte ebenso erzeugt werden kann wie eine auf Basis der absoluten oder relativen Häufigkeiten.

Neben dieser tabellarischen Ausgabe ist auch eine stärker visuelle Darstellung möglich. Dazu wird nach einem Rechtsklick auf das zu analysierende Dokument aus dem erscheinenden Menü der Eintrag *Wortwolke* gewählt. Das sich öffnende Fenster stellt alle gezählten Begriffe in alphabetischer Reihenfolge dar, wobei die Häufigkeit des jeweiligen Wortes durch seine Größe repräsentiert wird: Je größer ein Wort in der Wortwolke erscheint, desto häufiger ist es im analysierten Dokument. Somit ermöglicht die Wortwolke einen schnellen Einblick in die quantitativ dominierenden Begriffe eines Dokumentes. Allerdings ist es auch hier erforderlich, eine brauchbare Stoppliste zu erzeugen, was direkt im Fenster der Wortwolke über die Schaltfläche *Stoppliste editieren* geschehen kann. Wortwolken können für einzelne Dokumente erzeugt werden, aber auch für Dokumentgruppen, Dokumentsets oder den gesamten Datenkorpus auf einmal. Dazu ist jeweils auf die gewünschte Gruppe zu klicken, also etwa auf den Namen eines Sets oder auf den Eintrag *Dokumente* in der *Liste der Dokumente*.

Die Funktionen der *Worthäufigkeiten* und der *Wortwolke* stehen letztlich in der Tradition der quantitativen Inhaltsanalyse, wenngleich es sich dabei nur um einfachste Aspekte dieser Methode handelt. Die quantitative Inhaltsanalyse ist eine Verfahrensweise, die mithilfe statistischer Methoden *manifeste* Inhalte von Dokumenten untersucht. Ein übliches Verfahren ist dabei, Kategorien zu bilden, die Begriffe bündeln, denen inhaltlich die gleiche Bedeutung zugeschrieben wird. So könnten beispielsweise die Begriffe *Angst*, *Panik*, *Furcht*, *Beklemmung*, *Horror* und *Bammel* einer Kategorie mit dem Titel *Angst* zugeordnet werden. Die statistische Auswertung des zu analysierenden Materials erfolgt dann in den nächsten Schritten anhand eines Wortvergleichs: Indem Wort für Wort des Materials untersucht und mit den Begriffen der Kategorie *Angst* verglichen wird, wird gleichzeitig eine Auszählung vorgenommen, in der festgehalten wird, wie häufig die Kategorie *Angst* in Form der ihr zugeordneten Begriffe in jedem für die Analyse gewählten Dokument vorkommt. Am Ende dieses Prozesses liegt ein unmittelbarer quantitativer Überblick über die Bedeutung dieser Kategorie in jedem untersuchten Dokument vor. Ein weiteres plakatives Beispiel wäre das Vorhaben, die politische Ausrichtung eines Dokumentes zu analysieren. Hierfür könnten die Kategorien *Progressiv* und *Konservativ* angelegt werden, die jeweils

typische Begriffe in sich vereinen. Nach erfolgtem Auswertungsdurchlauf kann auf Grundlage *manifester* Inhalte eine Aussage bezüglich der politischen Ausrichtung des Dokuments getroffen werden. MAXQDA erlaubt mithilfe von MAXDictio eine Analyse in genau dieser Vorgehensweise. Den ersten Schritt dazu stellt die Definition von Kategorien im *Diktionär* dar. Das *Diktionär* stellt in MAXQDA die Gesamtheit der Kategorien dar. Jeder Kategorie kann eine beliebige Anzahl von Worten zugeordnet werden, Worte können auch mehreren Kategorien zugehörig sein. Um eine Analyse mit MAXDictio vornehmen zu können müssen demnach zunächst die Analysekategorien definiert werden, wozu der Eintrag *Diktionär* im Menü *MAXDictio* aufgerufen wird. Im sich öffnenden Fenster können die gewünschten Kategorien angelegt werden. Um den Kategorien Worte zuzuordnen, kann auf die bereits beschriebene Liste der *Worthäufigkeiten* zurückgegriffen werden. Um ein Wort aus dieser Liste einer Kategorie zuzuordnen, kann es einfach mit der Maus per *drag & drop* auf die entsprechende Kategorie gezogen werden. Soll ein Wort hingegen von der Zählung ausgeschlossen, also auf die *Stoppliste* gesetzt werden, so reicht es, in der ersten Spalte des Fensters *Worthäufigkeiten* einen Doppelklick auf den grünen Punkt auszuführen. Dieser wandelt sich daraufhin in einen roten Punkt, der anzeigt, dass das entsprechende Wort Teil der *Stoppliste* ist. Um nun die Auszählung zu starten wird die Funktion *MAXDictio-Codiertool* aus dem Menü *MAXDictio* aufgerufen. Als *Codieren* wird in diesem Kontext das Zählen der Häufigkeiten des Vorkommens der Kategorien bezeichnet, wobei die im *Diktionär* gesammelten Wörter als Indikatoren verwendet werden. Als Ergebnis wird eine Tabelle ausgegeben, die für jedes Dokument die Anzahl des Vorkommens jeder definierten Kategorie darstellt. Auch ist es möglich, das Dokument mithilfe von Codes weiter zu unterteilen, so dass sich die Auswertung nicht immer auf das gesamte Dokument beziehen muss, sondern auch feinere *Codiereinheiten* definiert werden können. Zudem ist es möglich, die Zählergebnisse direkt in *Dokumentvariablen* zu transformieren, wodurch sich wieder das gesamte Spektrum der oben beschriebenen Möglichkeiten zur Arbeit mit Variablen eröffnet.

Der Erstellung eines fundierten *Diktionärs* sowie einer guten Stoppliste kommt in dieser Forschungsmethodik eine entscheidende Rolle zu, da sie Analyseinstrumente bilden, die automatisch angewendet werden. Dabei bleibt ein gewisses Restrisiko, dass die Zählung nicht korrekt ist, etwa weil Wortteile als ganze Worte ausgewertet werden oder Begriffe mehrdeutig sind. Um die manuelle Überprüfung zu ermöglichen, kann MAXDictio eine *Validierungsdatei* erzeugen, welche den analysierten Text enthält und in der alle Zählungen kenntlich gemacht wurden, so dass die Grundlage der Auszählung transparent wird.

3.7 Forscher/innentriangulation: Intercoder-Übereinstimmung

Ein wichtiger Schritt zur Sicherung der Güte qualitativer Forschung, aber auch zur Validierung und Verbesserung der Forschungsergebnisse, ist der Ansatz, die Intersubjektivität des analytischen Vorgehens sicherzustellen (vgl. u.a. Seale 2007, Steinke 1999). Viele Möglichkeiten bietet hierzu die Arbeit in einem Team von Forscher/innen, wodurch unter anderem die von Norman Denzin unter dem Schlagwort *Forschertriangulation* (Denzin 1978) beschriebene Vorgehensweise nutzbar gemacht wird. Von besonderer Bedeutung kann dabei das Verfahren der Bestimmung der *Intercoder-Übereinstimmung* sein. In MAXQDA ist es möglich, die Codierungen, die zwei Forscher/innen unabhängig voneinander im gleichen Dokument vorgenommen haben, miteinander zu vergleichen. Ein wichtiges Ziel kann dabei sein, unterschiedliche Verständnisse von Dokument und/oder Code aufzudecken, die ggf. eine Redefinition und Verbesserung des Codesystems nach sich ziehen und zu zuverlässiger codierten – und damit interpretiertem, verstandenen – Material führen können. Dazu codieren mindestens zwei Personen unabhängig voneinander das gleiche Dokument, das zu diesem Zwecke zweimal in MAXQDA eingelesen wird und unterschiedlich benannt werden sollte, mit dem gleichen Codesystem. Anschließend wird über das Menü *Analyse* die Funktion *Intercoder-Übereinstimmung* aufgerufen. Im nun erscheinenden Dialog wird ausgewählt, welche Dokumente miteinander verglichen werden sollen. Zudem wird eingestellt, auf welche Übereinstimmung hin die Dokumente untersucht werden sollen. Zur Verfügung stehen hier drei Optionen. *Vorhandensein des Codes im Dokument* untersucht, ob in beiden Dokumenten die gleichen Codes angewendet wurden. Dabei spielt es keine Rolle, wie oft oder an welcher Stelle die Codes vergeben wurden, allein das Vorhandensein des Codes ist entscheidend. Daneben ist es möglich, nach der *Häufigkeit der Codes im Dokument* zu untersuchen. Hier wird auch berücksichtigt, ob beide Codierer/innen den fraglichen Code jeweils gleich häufig im Dokument vergeben haben, wobei die tatsächliche Position keine Rolle spielt. Schließlich ist es auch möglich, die *Übereinstimmung der Segmente in %* zu untersuchen. Diese Analyseeinstellung vergleicht, ob tatsächlich beide Codierende die identische Dokumentstelle dem gleichen Code zugewiesen haben. Natürlich wäre ein 100 prozentige Übereinstimmung eher zufällig, zumindest, wenn es sich um Material handelt, das in einigermaßen freier Form vorliegt. Deshalb ist es möglich, einen Prozentwert einzustellen, der als Grenze für gültige Übereinstimmungen fungiert. Er gibt an, welcher Anteil eines Codings mit jenem der anderen Codiererin/ des anderen Codierers übereinstimmen muss, um als übereinstimmend gewertet zu werden.

Neben der Möglichkeit, die Codierarbeit zweier verschiedener Personen miteinander zu vergleichen, kann diese Funktion auch dazu eingesetzt werden, die *Intracoder-Übereinstimmung* zu untersuchen. Dabei handelt es sich gewissermaßen um die Überprüfung der persönlichen Codierzuverlässigkeit. Dieses Verfahren kann sehr sinnvoll sein, um sicherzustellen, dass auch am Ende des Codierprozesses die Codes noch in der gleichen Weise vergeben – und damit verstanden – werden, wie am Anfang. Um die Intracoder-Übereinstimmung zu überprüfen, wird ein Dokument, das man in einer früheren Projektphase bereits codiert hat, erneut codiert, und zwar ohne sich die ursprünglichen Codierungen vorher anzusehen. Der weitere Verlauf ist dann identisch mit der beschriebenen Vorgehensweise zur Untersuchung der Intercoder-Übereinstimmung. Der einzige Unterschied liegt lediglich darin, dass als Dokumente, die verglichen werden sollen, die früher und die später von der gleichen Person codierten Versionen des identischen Dokuments gewählt werden.

Die Überprüfung der Intercoder-Übereinstimmung stellt eine wichtige Vorgehensweise der Triangulation dar, da auch hier eine zweite Perspektive auf das Datenmaterial eingenommen wird. Auch der Beitrag der Intracoder-Übereinstimmung ist nicht zu unterschätzen: Die Veränderungen, welche die Sicht auf das Forschungsthema, die Forschungsfrage, die befragten Personen etc. während des Forschungsprozesses erfahren, sind teilweise nicht unerheblich. Die Überprüfung der Konsistenz des Verständnisses und der praktischen Codierarbeit leistet somit ebenfalls einen direkten Beitrag zu Erhöhung der Qualität und hilft, ein sich verschiebendes Referenz- bzw. Deutungssystem zu erkennen und entsprechend zu reagieren.

4. Der schnelle Import gemischter Datenarten: Dokumente in Tabellen

Bisher wurde für die Integration quantitativer Daten in MAXQDA der Weg beschrieben, die betreffenden Daten zusätzlich zu den qualitativen separat in MAXQDA einzugeben oder mithilfe des Variablenimports einzulesen. In der Praxis gibt es jedoch auch einige Situationen, in denen die beiden Datenarten direkt gemeinsam in einer Datei vorliegen. Dies ist beispielsweise dann der Fall, wenn die Daten mithilfe eines Online-Datenerhebungstools oder unterstützt durch CATI-/CAPI-Software erhoben wurden. Aber auch nach der Eingabe oder Digitalisierung klassischer Papierfragebögen erhält man üblicherweise eine einzige Datentabelle. Dabei ist der Aufbau dieser Tabellen in der Regel gleich und orientiert sich an den Anforderungen von Software zur statistischen Datenanalyse, d.h. jede Zeile der Tabelle enthält einen Fall, jede Spalte steht für eine Frage, wobei sich Spalten mit standardisierten und nicht standardisierten Daten häufig abwechseln. Eine solche Urliste kann nach minimaler Vorbereitung direkt in MAXQDA eingelesen

werden, wobei gleichzeitig die unterschiedlichen Datenarten nach den Vorgaben der Forschenden getrennt und entweder als Teil eines Dokuments oder als Variablenwerte eingefügt werden, wobei auch hier eine Zeile als ein Fall gehandhabt wird. Beim Import werden für jede Zeile sämtliche Daten, die für die Übertragung in die *Liste der Dokumente* vorgesehen wurden – in der Regel die nicht standardisierten Daten – zusammengefasst, so dass je Zeile ein Dokument in der *Liste der Dokumente* entsteht, das all diese nicht standardisierten Daten enthält. Dabei wird für jede einzelne Zelle der Urliste, also letztlich für jeweils eine komplette Antwort auf eine einzelne Frage, ein eigener Absatz im Dokument erzeugt. Enthielte eine einzulesende Urliste also drei offene Fragen, so würde in MAXQDA jedes Dokument aus drei Absätzen bestehen. In der Urliste enthält in der Regel die erste Zeile als Spaltenüberschrift den Wortlaut der Fragen oder auch eine Fragenbezeichnung in Form eines Kürzels. MAXQDA erzeugt aus diesen Spaltenüberschriften der nicht standardisierten Daten zusätzlich automatisch Codes, denen als Codenamen eben diese Überschriften zugewiesen werden. Gleichzeitig werden alle Daten, die sich in der jeweiligen Spalte befanden, also die Antworten auf die entsprechende Frage darstellen, diesem neu erzeugten Code zugewiesen. So wird es möglich, direkt nach dem Import der Daten und ohne weiter manuelle Vorbereitung schnell auf einer explorativen Ebene mit der Analyse der Daten zu beginnen, beispielsweise, indem alle Antworten auf eine bestimmte Frage aufgerufen werden. Doch diese Funktion bringt noch mehr Komfort mit sich: Jene Daten, die nicht als auszuwertender Text, sondern als Variablenwerte zur Verfügung stehen sollen, werden automatisch in die *Dokumentvariablen* überführt. Dabei wird die Spaltenüberschrift, also der Fragewortlaut oder das Kürzel, zum Variablennamen und die von den befragten Personen angegebenen Daten werden direkt als Werte hinterlegt. Dadurch ergeben sich direkt nach dem Einlesen vielfältige Analysemöglichkeiten, viele der oben dargestellten Funktionen sind unmittelbar nutzbar. So können beispielsweise direkt Statistiken erzeugt werden, es können Kreuztabellen und Segmentmatrizen erzeugt werden oder es können mithilfe der *Aktivierung via Dokumentvariablen* die Antworten von Befragten mit bestimmten Merkmalen aufgerufen werden. Allerdings handelt es sich dabei in viele Fällen erst um den Beginn der Analyse, da die Codierung ausschließlich formal und nicht nach tatsächlichen inhaltlichen Kriterien erfolgte, eine interpretierende Bearbeitung des Materials also häufig noch notwendig sein wird. Insgesamt eröffnet diese Vorgehensweise mit vergleichsweise geringem Aufwand den Weg zu ernsthaften *mixed method*-Analysen der erhobenen Daten, in denen nicht, wie leider so oft, einer der beiden Datenarten eine viel zu geringe Beachtung zuteilwird.

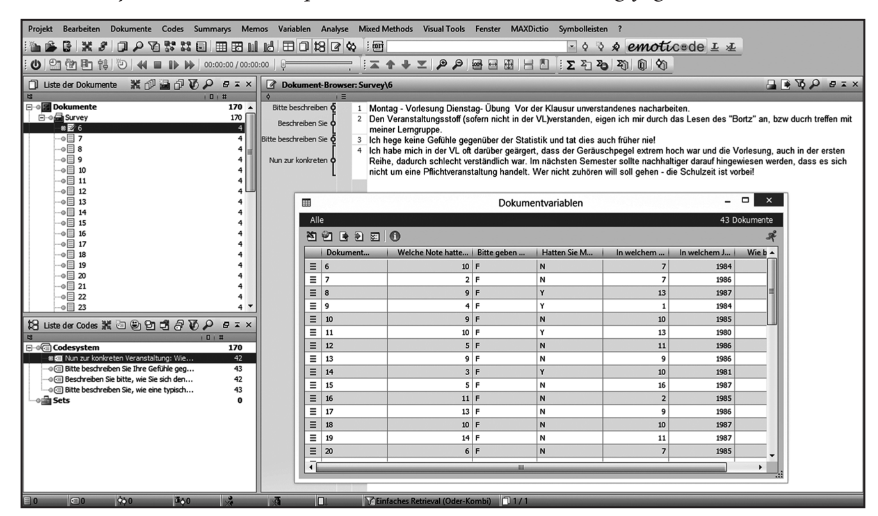
Um die Daten in MAXQDA einzulesen, muss zuerst mit der rechten Maustaste auf den Eintrag *Dokumente* in der *Liste der Dokumente* geklickt werden.

Im sich öffnenden Menü wird dann der Punkt *Import von Dokumenten aus Tabelle* aufgerufen. Im nächsten erscheinenden Fenster wird noch einmal erläutert, wie die Daten für den Import vorzubereiten sind. Die einzulesende Tabelle muss eine erste Spalte mit der Überschrift *Dokumentgruppe* enthalten, die in den Zellen darunter einen beliebigen Wert enthält. Hier wird definiert, welcher Dokumentgruppe der Datensatz der jeweiligen Zeile zugeordnet wird. Sollte die Dokumentgruppe nicht existieren, so wird sie erzeugt. Es können auch alle Dokumente in die gleiche Dokumentgruppe eingeordnet werden. Die zweite Spalte muss mit *Dokumentname* überschrieben werden. Der Wert, der hier hinterlegt wird, wird nach dem Import als Dokumentname in der *Liste der Dokumente* erscheinen. Es hat sich bewährt, hier einfach das Kürzel des Datensatzes, das ihm bei der Dateneingabe, während der Erhebung oder durch das Surveytool zugewiesen wurde, beizubehalten, nicht zuletzt, um die Verbindung zwischen den Dateien aufrecht zu erhalten. Die nachfolgenden Spalten enthalten schließlich die erhobenen Daten. Um MAXQDA mitzuteilen, welche der Spalten Werte enthalten, die als Variablen gespeichert werden sollen, müssen nur noch die Überschriften der entsprechenden Spalte bearbeitet werden: Jede Spalte, deren Überschrift mit einem *$*-Symbol beginnt, wird von MAXQDA als Variable behandelt. Um etwa die Werte der Spalte *Geburtsjahr* in die Variablenliste zu überführen würde die Überschrift der Spalte so bearbeitet, dass sie *$Geburtsjahr* heißt. Dadurch wird beim Import eine neue Variable des Namens *Geburtsjahr* erzeugt und die Werte werden dieser Variable zugewiesen. Alle übrigen Spalten werden als zu analysierender Text angesehen und als Absatz in das erzeugte Dokument eingefügt.

Abb. 15a: Ausschnitt einer fertig vorbereiteten Urliste: Die Spalten Dokumentgruppe *und* Dokumentname *existieren und sind gefüllt und die Überschriften der Spalten, deren Inhalt nicht als eigentlicher, zu codierender Text, sondern als Variablenwerte in MAXQDA eingefügt werden soll, beginnen mit einem* $*-Symbol.*

	A	B	C	D	E	F	G	H	I	J	K	L
1	Dokumentgru	Dokumentname	Bitte beschre	Beschreiben	Bitte beschre	Nun zur konk	$In welchem	$Bitte geben	$Welche Not	$Hatten Sie	$In welchem	$Wie bewert
2	Survey	6	Montag - Vo	Den Veranst	Ich hege keir	Ich habe mic	1984	F	10	N	7	2
3	Survey	7	Die Woche b	Für Montag a	Ich war nie s	In der Vorles	1986	F	2	N	7	2
4	Survey	8	Montag nach	mittels der Ü	Da Statistik r	Manchmal fie	1987	F	9	Y	13	2
5	Survey	9	Montag morg	Ich gehe zu	Zu Beginn de	Eigentlich ga	1984	F	4	Y	1	2
6	Survey	10	Montag 14-1	vor der vorle	Am anfang fa	ganz gut	1985	F	9	N	10	2
7	Survey	11	Vor jahreswe	Anfang auch	Vor Beginn se	Meist Kopfsc	1980	F	10	Y	13	2
8	Survey	12	montags geh	ich schreibe	kein gutes ge	oft überforde	1986	F	5	N	11	3
9	Survey	13	bis zum janu	ich hangel mi	ehrlich gesag	das tutorium	1986	F	9	N	9	2
10	Survey	14	Mo: Vorlesun	Tutorium, eig	ganz zu Begi	Vorlesung: ic	1981	F	3	Y	10	3
11	Survey	15	Die Statitikw	Ich höre mir	Ich war dem	Die Vorlesung	1987	F	5	N	16	2
12	Survey	16	montag morg	tutorium bes	zunächst wa	in den vorles	1985	F	11	N	2	2
13	Survey	17	Montag gehe	selber die ka	etwas ungut	Vorlesung: e	1986	F	13	N	9	3
14	Survey	18	Montag: Sta	1.: Ich höre	Ich konnte m	In der Vorles	1987	F	10	N	10	3
15	Survey	19	Montag mitta	Mir reicht die	Ich mochte i	Die Vorlesung	1987	F	14	N	11	2
16	Survey	20	Montag Mitta	Hauptsächlic	Anfangs hate	In der Vorleu	1985	F	6	N	7	2
17	Survey	21	Montag Morg	Nach der Vor	Nein Statistik	Gut. Teilweis	1986	F	2	N	15	2
18	Survey	22	Montag 14.0	durch Übung	Anfangs hatt	Vorlesung: m	1986	F	7	N	10	4
19	Survey	23	Meine Statist	Ich erarbeite	Ich finde, da	Vorlesung: E	1986	F	10	N	15	3
20	Survey	24	Montag Vorle	Indem ich de	Ja ein bissch	Meistens gar	1987	F	3	N	16	3
21	Survey	25	Meine typiscl	Vorlesung un	Das Thema a	Alle Veransta	1986	M	5	Y	9	2
22	Survey	26	Eine typische	Durch Nachb	Durch Erzähl	Allgemein seh	1985	F	7	N	5	1
23	Survey	27	Vor der Stati	#NAME?	Ich hatte et	Ich habe mic	1983	M	7	N	11	2
24	Survey	28	Montags Vor	Indem ich die	Meine gefühl	Mit Hilfe der	1985	F	10	N	10	2
25	Survey	29	Meine Statist	Also zum Ein	Zu Beginn, m	Ja in der Vorl	1986	F	3	N	10	2
26	Survey	30	Ich besuche	siehe oben	Zu Beginn eher Angst und		1986	F	6	N	7	2

Abb. 15b: Ergebnis des Imports der vorbereiteten Urliste: Aus jeder Zeile der Urliste wurde ein eigenes Dokument, in dem jede Zelle als Absatz erscheint. Die als Variablen markierten Werte ($-Symbol) wurden entsprechend in die Variablenliste eingefügt.



5. Resümee: Triangulation mithilfe von QDA-Software

Die Verbindung qualitativer und quantitativer Daten und Forschungslogiken entwickelt sich unabhängig von der konkreten Bezeichnung der jeweils angewendeten Vorgehensweise mehr und mehr zum Standard sozialwissenschaftlicher Forschung. Moderne QDA-Software kann hier auf viele Weisen unterstützen. Durch die Möglichkeit, Daten unterschiedlicher Art gemeinsam zu betrachten, gleichzeitig auszuwerten und auf unterschiedlichste Weise aufeinander zu beziehen eröffnet sich ein weites Spektrum analytischer Möglichkeiten, das vor Zeiten der computerunterstützten Datenanalyse fast undenkbar war. Arbeitsschritte, die manuell einen enormen Zeitaufwand bedeutet hätten, sind mithilfe des Computers in wenigen Augenblicken abgeschlossen – man denke etwa an die sich durch die *Aktivierung via Dokumentvariablen* oder durch *Joint Displays* ergebenden Möglichkeiten und die Arbeit, die eine manuelle Darstellung bedeuten würde. Von gleicher Wichtigkeit ist die Öffnung der Software für die Interaktion mit anderen, spezialisierten Programmen, hier vor allem mit Werkzeugen zur statistischen Datenanalyse. In der Gesamtheit betrachtet ist es für die Forschenden heute auf technischer Ebene leichter als jemals zuvor, *echte Mixed Methods*-Analysen durchzuführen, die über

das bloße Nebeneinanderstellen unterschiedlicher Daten(-arten) hinausgehen. Gleichzeitig wird die Erfordernis fundierter methodologischer Kenntnisse immer dringlicher, da Forschende nicht nur in beiden Analysearten bewandert sein müssen, sondern auch die durch die Triangulation entstehende Komplexität zu bewältigen haben. Schließlich ist nicht alles, was sich transformieren, mixen, triangulieren oder integrieren lässt, auch wirklich sinnvoll. Dabei ist nicht aus den Augen zu verlieren, dass es sich bei diesen Ansätzen nicht um Selbstzwecke handelt, sondern dass eine geschickte, auf die Besonderheiten der jeweiligen Untersuchung zugeschnittene Verbindung der Methodenfamilien tatsächlich einen enormen Zugewinn an Qualität und Güte, Verlässlichkeit und Glaubwürdigkeit mit sich bringen kann (siehe etwa Kelle 2008, Flick 2011, Mayring 2001).

Literatur

Creswell, John W. / Plana-Clark, Vicki L. (2011). *Designing and Conducting Mixed Methods Research*. 2. Aufl. Thousand Oaks u.a.: Sage Publications.

Corbin, Juliet / Strauss, Anselm (2008). *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*. 3. Aufl. Thousand Oaks u.a.: Sage Publications.

Denzin, Norman K. (1978). *The Research Act. A Theoretical Introduction to Sociological Methods*. 2. Aufl. New York u.a.: McGraw-Hill.

Flick, Uwe (2011). *Triangulation. Eine Einführung*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kelle, Udo (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kuckartz, Udo (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim und Basel: Beltz Juventa.

Kuckartz, Udo (2010). *Einführung in die computergestützte Analyse qualitativer Daten*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Mayring, Philipp (2001). Kombination und Integration qualitativer und quantitativer Analyse. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* 2(1), Art. 6, Online: http://nbn-resolving.de/urn:nbn:de:0114-fqs010162 [26.06.14]

Ragin, Charles C. (2000). *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Seale, Clive (1999). *The Quality of Qualitative Research*. London u.a.: Sage Publications.

Settinieri, Julia (2014). Forschst Du noch, oder triangulierst Du schon? In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 17-35.

Steinke, Ines (1999). *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Weinheim u.a.: Juventa.

Strübing, Jörg (2008). *Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

“Are you good at reading?” Zur Diagnosekompetenz von Englischlehrkräften

Laura Armbrust

Das hier beschriebene Dissertationsvorhaben untersucht empirisch die Diagnosekompetenz von Englischlehrkräften der Sekundarstufe I hinsichtlich der Fertigkeit Lesen. Grundlage der Datenerhebung bilden leitfadengestützte Experteninterviews. Der vorliegende Beitrag fokussiert insbesondere die Auswertungsmethode der *Qualitativen Inhaltsanalyse*. Zunächst wird über einen theoretischen Hintergrund in das Thema des Forschungsprojekts eingeführt. Hieran schließen sich Forschungsdesign sowie Erläuterungen zur Datenerhebungsphase an. Ausführlich werden im Anschluss Methode und Methodik der *Zusammenfassenden Inhaltsanalyse* und *Inhaltlich Strukturierender Inhaltsanalyse* veranschaulicht und die Vorzüge einer computerunterstützten Auswertung mithilfe von MAXQDA herausgearbeitet.

1. Was wir über das Diagnostizieren von Lesekompetenz wissen

Schon in den 1990er Jahren unterstrich Ingenkamp (1991: 780) die Relevanz von Diagnosekompetenz: „Ein Bildungswesen ohne entwickelte Pädagogische Diagnostik ist ein lernunfähiges System.“ Dass sich hieran auch gut 20 Jahre später nichts geändert hat, zeigen beispielsweise Paradies et al. (2011: 10), die beklagen, dass aufgrund einer unzureichenden Diagnosekompetenz auf Seiten der Lehrkräfte Lernschwierigkeiten der Schülerinnen und Schüler häufig unerkannt blieben. Die Forderung nach einer Verbesserung der Diagnosekompetenz von Lehrkräften ist spätestens seit dem unbefriedigenden Abschneiden deutscher Schülerinnen und Schüler in den (internationalen) Vergleichsstudien PISA, TIMSS, oder IGLU in das Bewusstsein der Öffentlichkeit und in den Fokus bildungspolitischer Diskussionen gerückt. So wurden Hauptschullehrkräfte in PISA 2001 gebeten, sogenannte „schwache Leser“ (vgl. Hesse/Latzko 2011: 14) zu identifizieren. Die Einschätzungen der Lehrkräfte wurden anschließend mit dem tatsächlichen Abschneiden der Schülerinnen und Schüler im Bereich Lesekompetenz verglichen. Knapp 90% derjenigen Lernenden, die im Test als schwache Leser eingestuft wurden, konnten von den Lehrkräften vorab nicht als solche identifiziert werden (vgl. ibid.: 14f.).

Die Kultusministerkonferenz der Länder (KMK) reagierte 2003 und 2004 auf das enttäuschende Ergebnis der Vergleichsstudien mit der Abfassung von

Bildungsstandards für die Fächer Mathematik, Deutsch und die erste Fremdsprache (Englisch/Französisch). Darüber hinaus wurden Standards für die Lehreraus- und -fortbildung formuliert. Diese beziehen explizit auch die Diagnosekompetenz mit ein: „Lehrerinnen und Lehrer üben ihre Beurteilungs- und Beratungsaufgabe im Unterricht und bei der Vergabe von Berechtigungen für Ausbildungs- und Berufswege kompetent, gerecht und verantwortungsbewusst aus. Dafür sind hohe pädagogisch-psychologische und diagnostische Kompetenzen von Lehrkräften erforderlich." (KMK 2004: 3). Diese diagnostischen Kompetenzen gilt es, differenziert zu betrachten und gezielt einzusetzen, sodass Lehrkräfte gegenstandsangemessen urteilen und verlässlich diagnostizieren (vgl. KMK 2004, Schrader 2009). Im Einzelnen bedeutet dies, dass Lehrkräfte in der Lage sein sollten, auf das Leseverständnis als Solches zu achten und bei dessen Diagnose zum Beispiel mögliche fehlende grammatisch-lexikalische Kompetenzen der Schülerinnen und Schüler außer Acht zu lassen. Diagnosekompetenz wird damit als herausfordernder und wichtiger Bestandteil der unterrichtlichen Tuns der Lehrkräfte anerkannt.

Nicht nur auf bildungspolitischer Ebene ist die Diagnosekompetenz explizit in die Lehrerbildung aufgenommen worden (vgl. KMK 2004), zahlreiche Studien beschäftigen sich ebenfalls mit diesem Thema. Sie untersuchen in erster Linie die Urteilsgenauigkeit von Lehrkräften in unterschiedlichen Fachdisziplinen (u.a. Spinath 2005, Südkamp et al. 2008, McElvany et al. 2009) und bringen durchaus unterschiedliche Ergebnisse hervor. So untersuchte beispielsweise Spinath (2005) eine Gruppe von 723 Grundschulkindern in Bezug auf ihre Intelligenzkennwerte sowie Selbsteinschätzungen hinsichtlich ihrer Fähigkeitswahrnehmung, Lernmotivation und Schulangst. Diese Daten wurden mit den Aussagen von 43 Klassenlehrkräften korreliert, die ebenfalls die Schülermerkmale eingeschätzt hatten. Das Ergebnis der Studie wies auf eine geringe Akkuratheit der Einschätzungen der Lehrkräfte hinsichtlich der Schülermerkmale hin (vgl. Spinath 2005: 85ff.). Die Studie von Südkamp et al. (2008) untersuchte die diagnostische Kompetenz von 24 Lehramtsstudierenden, die mittels Computersimulation eines Klassenzimmers Leistungen von Schülerinnen und Schülern möglichst exakt einschätzen sollten. Bei zwei zeitlich versetzten Messzeitpunkten erwiesen sich die Einschätzungen der Lehramtsstudierenden als „recht gute diagnostische Leistungen" (Südkamp et al. 2008: 95). Diese experimentelle Untersuchung bildet jedoch auf Grund des eingesetzten simulierten Klassenraums keine komplex ablaufenden Prozesse der Unterrichtswirklichkeit ab.

Zentrales Ergebnis der Studie von McElvany et al. (2009), in der 116 Lehrkräfte und 48 Klassen untersucht wurden, ist eine schwache bis mittelmäßig ausgeprägte Diagnosekompetenz der Lehrkräfte mit Tendenz zur Unterschätzung

der Schülerleistungen. Lehrkräfte wurden hierfür gebeten, Schülerleistungen und Aufgabenschwierigkeiten anhand von Lernmedien mit Abbildungen einzuschätzen. Die Ergebnisse zeigen Schwierigkeiten für die Lehrpraxis aus folgendem Grund: Das Wissen der Lehrkräfte hinsichtlich der Bedeutung von Text und Abbildung ist nicht hinreichend. Dies ist insbesondere vor dem Hintergrund der Lesekompetenzförderung bemerkenswert, da Bilder als Verstehenshilfe im Leseprozess gesehen werden können (vgl. u.a. BMBF 2007: 24ff.).

Während die bisher dargestellten Studien in ihren Vorgehensweisen dem quantitativen Paradigma zuzuordnen sind, ist nur eine qualitativ angelegte Studie bekannt, welche sich der Frage nach Beurteilerübereinstimmungen im Kontext des Englischunterrichts der Grundschule widmet (Hochstetter 2011). Hierzu wurden Beobachtungsbögen eingesetzt, um die Fertigkeit Sprechen bei Grundschulkindern zu diagnostizieren. Hochstetter attestiert den beteiligten Lehrkräften eine zufriedenstellende Übereinstimmung der jeweiligen Einschätzungen in Bezug auf die Schülerleistungen. Im Gegensatz dazu beschreibt sie jedoch auch Unsicherheiten der Befragten insbesondere hinsichtlich der vorläufigen Diagnose von Schülerleistungen, die nicht für die Notengewinnung bzw. zur Leistungsmessung dienen (vgl. Hochstetter 2011: 244f.).

Durch Studien bislang nicht in den Blick genommen wurden hingegen Diagnosekompetenzen bezüglich der rezeptiven Fertigkeiten, insbesondere des Lesens, das im Sinne einer gesellschaftlichen Teilhabe an Bildung als Schlüsselkompetenz gelten kann. In PISA 2000 und 2009 lagen die deutschsprachigen Leseleistungen der Schülerinnen und Schüler in Deutschland unter dem OECD-Durchschnitt und damit hinter den Erwartungen der Bildungspolitik. Ebenso zeigten die DESI-Studie 2006 und der Ländervergleich 2009 deutliche Schwächen in der englischen Lesekompetenzleistung deutscher Schülerinnen und Schüler (vgl. Harsch et al. 2006, Köller et al. 2010). Die ungenügende Leseleistung beider Schulstufen in den Fächern Deutsch und Englisch v.a. in den nicht-gymnasialen Schulformen lässt deutlich werden, dass die Förderung von Lesekompetenz und damit einhergehend die Fähigkeit der Lehrkräfte, Lesekompetenz exakt zu diagnostizieren, als ein wichtiges Desiderat angesehen werden muss.

Zweifelsohne kann nicht nur alleine der Deutschunterricht gefordert sein, sondern es muss vielmehr Aufgabe aller Fächer sein, Lesekompetenz zu fördern, da sie als fächerübergreifende Schlüsselkompetenz gilt (vgl. Spinner 2004: 136 u. 138). Naheliegend ist die Fokussierung der Lesekompetenz auch in den Fremdsprachen, insbesondere im Fach Englisch. In diesem Fach wird viel gelesen und mit Texten gearbeitet, sodass die Schülerinnen und Schüler über eine entsprechende Lesekompetenz verfügen müssen. Es liegt also nahe, der Lesekompetenzförderung verstärkt Beachtung zu schenken, da ab der Sekundarstufe

I insbesondere die Schriftsprache als wichtigste Kompetenz in allen Fächern gefordert wird. Um kompetente Leserinnen und Leser auszubilden, bedarf es einer sachgerechten Diagnose von Seiten der Lehrkräfte (vgl. Schrader 2009). Lesekompetenzdiagnose und eine sich anschließende Förderung sind entscheidende Faktoren im Unterricht, die sich nach Einführung des frühen Fremdsprachenlernens in der Primarstufe in allen Bundesländern (2003) auch in den ersten beiden Jahren der Sekundarstufe I von Bedeutung ist. Die Schülerinnen und Schüler haben bereits erste Kenntnisse in der Fremdsprache während der Primarstufenzeit erworben, die sich jedoch weitestgehend auf die Mündlichkeit konzentrieren. In den vergangenen Jahren verlagerte sich der Fokus von der reinen Mündlichkeit hin zu einem auch an der Schriftsprache orientierten Ansatz, sodass zu Beginn der Sekundarstufe I auf erste Kenntnisse in allen Kompetenzbereichen der Fremdsprache zurückgegriffen werden kann (vgl. u.a. Hessisches Kultusministerium o.A.: 11). Dieses Potenzial gilt es für den jeweiligen Englischunterricht zu Beginn der Sekundarstufe I zu nutzen und im Sinne einer gegenstandsangemessenen Diagnose von Seiten der Lehrkräfte aktiv in pädagogische Überlegungen mit einzubeziehen. Daher müssen oftmals unbewusst ablaufende Prozesse der Diagnose (vgl. Hochstetter 2011) in das Bewusstsein der Lehrkräfte gerückt werden, um einen bedarfsgerechten und motivierenden englischen Leseunterricht gestalten zu können, Kompetenzzuwächse zu ermöglichen bzw. Defizite frühzeitig zu erkennen und auszugleichen. So ergibt sich im Sinne Hallets (2006: 101) ein Ziel, „welches letztlich eine diagnostische Kompetenz bei Lehrer/innen sein [muss], die sie in die Lage versetzt, routinemäßig Diagnoseinstrumente selbst zu entwickeln und diese bedarfsorientiert einzusetzen."

Vor diesem Hintergrund liegt der Fokus des hier beschriebenen Dissertationsvorhaben auf der Erforschung der Diagnosekompetenz bei Englischlehrkräften der Sekundarstufe I in Bezug auf fremdsprachliches Lesen. Dabei stehen folgende Fragen im Mittelpunkt:

1. Wie diagnostizieren Lehrkräfte fremdsprachliche Lesekompetenz und welche Methoden wenden sie dabei an?
2. Welche Bedingungsfaktoren der Diagnosekompetenz (Evaluierungskompetenz, Kompetenzwissen, Bedingungswissen, technologisches Wissen oder diagnostische Methodenkompetenz, Änderungswissen und Vergleichswissen) zeichnen die Lehrkräfte im Bereich fremdsprachliche Lesekompetenz aus und wie stellen sich diese dar?
3. Welcher Maßnahmen (z. B. Fortbildung zum Einsatz von Diagnoseinstrumenten) bedarf es, um Lehrkräfte im Bereich der Lesekompetenzdiagnose zu professionalisieren?

2. Forschungsdesign

Dem Kriterium der Gegenstandsangemessenheit folgend wurden zur Datenerhebung leitfadengestützte Experteninterviews eingesetzt (vgl. u.a. Bogner et al. 2005, Schrader 2009: 239, Gläser/Laudel 2010, Meuser/Nagel 2010, Kruse 2011: 255ff.). Diese haben sich mit Blick auf die Fragestellung als zielführend erwiesen, um das exklusive (Erfahrungs-) Wissen der Lehrkräfte hinsichtlich der Diagnose von fremdsprachiger Lesekompetenz zu eruieren. Für die hier beschriebene Arbeit relevante Typologien des Experteninterviews beinhalten das ‚systematisierende' und ‚theoriegenerierende' Interview als „eine anwendungsbezogene Variante von Leitfadeninterviews" (Kruse 2011: 57). Im Vordergrund des systematisierenden Experteninterviews steht „das aus der Praxis gewonnene, reflexiv verfügbare und spontan kommunizierbare Handlungs- und Erfahrungswissen" (Bogner/Menz 2005: 37). Das theoriegenerierende Experteninterview hingegen strebt eine analytische Rekonstruktion von implizitem und damit unbewusstem Deutungswissen sowie Handlungsroutinen der Lehrkräfte an. Beiden Dimensionen gemein ist deren Bestreben nach thematischer Vergleichbarkeit der Expertenäußerungen. Ihre Relevanz für das Forschungsvorhaben erhalten sie aus unterschiedlichen Motivationen. So ist es einerseits entscheidend, systematisch (Experten-) Wissen auf Basis des fachdidaktischen Wissens über Diagnose zu eruieren und im Kontext des eigenen fremdsprachlichen Leseunterrichts zu beleuchten sowie das generelle theoretische Wissen über Diagnoseinstrumente zu erforschen. Andererseits sind implizite Wissens- und Deutungsmodelle der Lehrkräfte in Bezug auf die „subjektive Dimension" (Bogner/Menz 2005: 38) als persönliche Einschätzung der Leseleistungen der Schülerinnen von Bedeutung für die Forschungsfrage. Den Status „Experten" erhielten die Lehrkräfte aufgrund ihres professionellen Handlungsfeldes sowie angesichts ihres speziellen Wissens und ihrer Expertise in fachdidaktischen und fachwissenschaftlichen Bereichen.[1]

Die leitfadengestützten Interviews wurden als Einzelinterviews geführt, um individuelle Diagnosemaßnahmen und spezifisches unterrichtliches Handeln zu erforschen. Der Leitfaden bewegte sich auf einem Kontinuum von starker Strukturierung des Interviews bis hin zu einer offenen Gesprächsführung mit abwechselnd stark geführten und offenen Fragen. So ergaben sich für die Interviewführung zwei Phasen: In der ersten Phase wurden die Lehrkräfte gebeten,

1 Zur Unterscheidung der Dimensionen von Expertenwissen siehe u.a. Bogner/Menz (2005: 43f.) und Meuser/Nagel (2005: 73ff.).

zwei kurze Videosequenzen (*prompts*[2]) von Schülerinnen der fünften Klassenstufe anzuschauen. Die videografierten Schülerinnen lasen einzeln einen kurzen fremdsprachigen Sachtext und wurden dazu aufgefordert, Lesestrategien zu verwenden, die für die Klassenstufe als relevant angesehen werden können, z.B. Wichtiges unterstreichen, *intelligent guessing*, etc. Die Handlungen der Schülerinnen und Schüler ermöglichten eine direkte Beobachtung, ob und welchem Umfang Lesestrategien angewendet werden. Fragen über das individuelle Leseverhalten, die Lesefreude und die Textsortenpräferenz der Schülerinnen wurden ebenfalls in das Video mit aufgenommen. Im Anschluss an die Betrachtung der Videoausschnitte sollten die Lehrkräfte die fremdsprachlichen Leseleistungen der Schülerinnen und Schüler (subjektiv) einschätzen. Dazu wurden ihnen sehr offene Fragen gestellt (z.B. Wie schätzen Sie die Leseleistung ein und wie würden Sie Ihre Einschätzung begründen?). In der zweiten Phase des Interviews wurden stärker strukturierte Fragen zu fachdidaktischem Wissen von Diagnose im fremdsprachlichen Leseunterricht, zu theoretischem Wissen über Diagnoseinstrumente sowie zu den Fortbildungsbedürfnissen der Lehrkräfte hinsichtlich der eigenen Diagnosekompetenz gestellt. Zentrale Ziele der zweiten Phase des Interviews waren, das Erfahrungswissen der Lehrkräfte sowie das Wissen über theoretische Konstrukte von Diagnose(-kompetenz) zu erfragen. So leitete eine erste Frage (Wie würden Sie das Lesen bei den einzelnen Schülerinnen einschätzen?) offen in die Interviewsituation ein und ermöglichte den Lehrkräften von der Situation ausgehend zu beschreiben, was sie sehen und dies mit ihrem (Fach-) Wissen zu verknüpfen.

Das Datenmaterial aus der ersten Phase der Interviews wurde also mithilfe einer Erzählaufforderung (ausgelöst durch einen visuellen Impuls) generiert. In der zweiten Phase des Interviews konnte das in der Erzählung explizierte, ursprünglich implizite Wissen der Lehrkräfte über Diagnose von fremdsprachlicher Lesekompetenz mit explizit verfügbarem Wissen aus der individuellen Lehrpraxis (in Form verbaler Daten) in Beziehung gesetzt und trianguliert werden. Die Triangulation zweier Datensätze (Erzählaufforderung und Leitfadeninterview) innerhalb einer Methode (Interview) lässt sich der Methodentriangulation und hier insbesondere der *within-method triangulation* zuordnen (vgl. Flick 2011: 27f.). Das Verständnis von Triangulation, das dem hier beschriebenen Vorhaben zugrunde liegt, lässt sich als Erkenntniserweiterung bezeichnen, wobei die

---

2 In Anlehnung an das Erhebungsinstrument *stimulated recal*l (vgl. u. a. Gass/Mackey 2000) verwendetes Instrument, das die Interviewpartner zu Kommentaren bzgl. der vorgespielten Situation veranlassen soll.

einzelnen Datensätze eine komplementäre Funktion haben und nicht notwendigerweise Konvergenz anstreben, wie dies im Falle der Triangulation als Validierungsstrategie erstrebenswert ist (vgl. Settinieri in diesem Band). Durch diese Konstruktion der Interviewsituation erhoffte ich mir eine vielschichtige und multiperspektivische Ergänzung meines Forschungsgegenstandes: „We should combine theories and methods carefully and purposefully with the intention of adding breadth or depth to our analysis but not for the purpose of pursuing ‚objective' truth" (Fielding/Fielding 1986: 33).

3. Datenerhebung: Planung und Durchführung der Interviews

Für die Datenerhebung wurden dreizehn Personen mittels des ‚Schneeballprinzips' (vgl. Kruse 2011: 89) ausgewählt. Hierzu sprach ich eine mir bekannte Lehrerin an, die mich dann auf weitere Kolleginnen und Kollegen verwies, die wiederum weitere Lehrkräfte kannten, etc. Die insgesamt zehn weiblichen und drei männlichen Lehrkräfte unterrichten am Gymnasium (neun Lehrkräfte), an einer Integrierten Gesamtschule (zwei Lehrkräfte), und an einer Kooperativen Gesamtschule (eine Lehrkraft) im Großraum Frankfurt am Main. Die Interviews wurden allesamt an den jeweiligen Schulen der Lehrkräfte durchgeführt und erforderten ein hohes Maß an Flexibilität auf Seiten der Forscherin (z.B. unterschiedliche Zeiten an unterschiedlichen Tagen, Zeitdruck der Lehrkräfte auf Grund sich anschließender Unterrichtsstunden, Störgeräusche, etc.). Die Befragungen dauerten durchschnittlich 60 Minuten.

Ausschlaggebende Kriterien für die Auswahl der Interviewpersonen waren das Abhalten des Englischunterrichts in Klasse fünf/sechs, die zu unterrichtende Schulart sowie eine breite Streuung der Studienabschlüsse des Zweiten Staatsexamens (vor und nach der bildungspolitischen Intervention der KMK und der Einführung der Bildungsstandards) und die damit verbundene Berufserfahrung. Ziel war es, die implizite und explizite Diagnosekompetenz von Lehrkräften aller Schularten möglichst umfassend abzubilden. Von Interesse war zudem die Frage, ob sich Unterschiede in der jeweiligen Diagnosekompetenz der Befragten mit Blick auf ihre Berufserfahrung bzw. den Zeitpunkt des Studienabschlusses feststellen lassen? Mithilfe dieser Kriterien sollte eine möglichst große Perspektivenvielfalt gewährleistet werden.

4. Der Auswertungsprozess

Nach der erfolgreichen Phase der Datenerhebung wurden die Interviews transkribiert und für den Auswertungsprozess aufbereitet. Kvales (2009) Anspruch

für eine umsichtige Datenauswertung – „entering into a dialogue with the text, going into an imagined conversation with the ‚author' about the meaning of the text" (Kvale 2009: 192) – stellt zwei Leitprinzipien vor, die für die Wahl der Auswertungsmethode des hier beschriebenen Forschungsvorhabens ebenfalls von Bedeutung sind: Offenheit und das Arbeiten am Datenmaterial. Das Prinzip der Offenheit (vgl. u.a. Steinke 1999: 35ff.) wird in der Qualitativen Sozialforschung oftmals beschrieben als starke Zurückhaltung des Forschenden gegenüber theoretisch-geleiteten Deutungen des Materials (vgl. Lamnek 2002: 165). Dieses Prinzip kommt in der Wahl einer induktiven Auswertungsmethode – der *Zusammenfassenden Inhaltsanalyse* – zur Anwendung, die aber dennoch von einer gewissen „theoretischen Sensibilität" (Glaser/Strauss 1998: 54) gekennzeichnet ist (z.B. durch Kenntnisse aus Literaturrecherche).

Angesichts der methodischen Vorüberlegungen zum systematisierenden und theoriegenerierenden Experteninterview erweist sich die *Qualitative Inhaltsanalyse* (vgl. Mayring 2008, Gläser/Laudel 2010, Kuckartz 2012[3]) als sinnvoll, die folgende Merkmale aufweist: „Die Qualitative Inhaltsanalyse will sprachliches Material systematisch analysieren, indem sie das Material zergliedert und schrittweise bearbeitet, theoriegeleitet am Material ein Kategoriensystem entwickelt und die Analyseaspekte vorher festlegt" (Kuckartz 2010: 92f.). Die *Qualitative Inhaltsanalyse* ist also ein regelgeleitetes und systematisches Verfahren, dessen Intention es ist, „das Material so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben, durch Abstraktion einen überschaubaren Corpus zu schaffen, der immer noch Abbild des Grundmaterials ist" (Mayring 2008: 58).

In einem ersten Analyseschritt werden die Interviews mittels der *Zusammenfassenden Inhaltsanalyse* ausgewertet (vgl. Mayring 2008: 59ff.), um das Wissen der befragten Experten möglichst genau und unvoreingenommen zu beschreiben. Daher werden zunächst induktiv Kategorien gebildet mit dem Ziel, die komplexen inhaltlichen Aussagen der Lehrkräfte abzubilden und um eine möglichst detailgetreue Abbildung vom Datenmaterial, ähnlich dem Offenen Codieren in der *Grounded Theory* (Glaser/Strauss 1998), zu erhalten. In einem zweiten Schritt wird die Zusammenfassende Inhaltsanalyse in die *Inhaltlich Strukturierende*

3 Die hier benannten Autoren bedingen und ergänzen sich gegenseitig. So hat Mayring in seinem Werk sein Verfahren der qualitativen Inhaltsanalyse vertieft, welches Kuckartz (2012) aufgreift und insbesondere vor dem Hintergrund einer computergestützten Auswertung beleuchtet. Gläser/Laudel (2010) wiederum übertragen die Auswertungsmethode Mayrings gezielt auf das Experteninterview, zeigen Vorzüge und Nachteile dieser Methode auf.

Inhaltsanalyse integriert, um die aus den Bezugstheorien (Diagnose und fremdsprachliches Lesen) stammenden Kategorien deduktiv, gleichsam als Schablone auf das Material zu projizieren. Die alleinige Verwendung der *Zusammenfassenden Inhaltsanalyse* erscheint mir für die vorliegende Forschungsarbeit und die damit verbundene Forschungsfrage nicht zielführend, da das Gütekriterium „intersubjektive Nachvollziehbarkeit" (Steinke 2007: 324ff.) hinsichtlich der thematischen Vergleichbarkeit der Aussagen nicht gegenstandsangemessen gesichert werden konnte. Mit der Integration *Zusammenfassenden Inhaltsanalyse* in die *Inhaltlich Strukturierte Inhaltsanalyse* kann gewährleistet werden, dass theoriegeleitete Kategorien das induktiv gebildete Kategoriensystem ergänzen und der Auswertung eine größere „Differenziertheit, Komplexität und Erklärungskraft" (Kuckartz 2012: 77) verleihen.

Die nachfolgende Abbildung soll das methodische Vorgehen des Auswertungsprozess veranschaulichen:

Abb. 1: Methodisches Vorgehen (in Anlehnung an Kuckartz 2012).

Schritt 1:
Anwenden der Hauptkategorien aus Leitfaden bei gleichzeitiger sequentieller Paraphrasierung des gesamten Datenmaterials mit MAXQDA-Software

Schritt 2:
Generalisierung der gesamten Codiereinheiten des Datenmaterials auf festgelegtes Abstraktionsniveau

Schritt 3:
Streichen und Bündeln von bedeutungsgleichen Codiereinheiten mittels Text-Retrieval

Schritt 4:
induktives Bilden von Sub-Kategorien am gesamten Datenmaterial entlang

Schritt 5:
konsensuelles Codieren mit Forschungspartner (Forschertriangulation)

Schritt 6:
synoptische und querschnittsartige Zusammenstellung des Kategoriensystems als Text-Retrieval

Schritt 7:
Anwenden der deduktiven Kategorien aus Bezugstheorien auf bestehendes Kategoriensystem

Schritt 8:
kategorienbasierte Auswertung an Hauptthemen (inter-und intra-textuell) sowie Ergebnisdarstellung

Forschungsfrage

Schritt 1:
Ausgehend von der Forschungsfrage und dem Leitfaden ergeben sich folgende vorläufige thematische Hauptkategorien, die als Raster zur Gliederung der Fülle des Datenmaterials im Sinne von Analyseheuristiken eingesetzt werden: ‚Einschätzung der videografierten Schülerinnen', ‚Unterrichtsmethoden', ‚generelle Verfahren zur Diagnose', ‚Kenntnis über Diagnoseinstrumente' und ‚Wünsche für die Weiterbildung'. Auf Grundlage dieser Hauptkategorien werden das gesamte Datenmaterial sequenziell analysiert und relevante Sinneinheiten durch Paraphrasierung den Kategorien zugeordnet. Unter Sinneinheiten sind mindestens ein in sich verständlicher Teilsatz bis hin zu bedeutsamen Absätzen einschließlich der Interview-Frage zu verstehen, falls diese für die Interpretation der Aussage relevant ist. „Wichtigstes Kriterium ist, dass die Textstelle ohne den sie umgebenden Text für sich alleine ausreichend verständlich ist" (Kuckartz 2012: 82). Zudem werden Textstellen mit derselben thematischen Information ebenfalls paraphrasiert, um einer vorschnellen Interpretation vorzubeugen. Bei der Analyse fiel auf, dass die Befragten immer wieder weitere Themen anschnitten, wie zum Beispiel Übergang von Primar- zur Sekundarstufe, Differenzierung oder bildungs- und schulpolitische Themen, sodass das Kategoriensystem um zwei weitere Hauptkategorien zu ergänzen war: ‚Allgemeine Aussagen über Unterricht und Lernende' (ähnlich dem *offenen Codieren* in der *Grounded Theory*, vgl. Glaser/Strauss 1998) und ‚Bildungspolitik/Schulpolitik'. Diese Erweiterung der Hauptkategorien macht deutlich, dass die Kontexte der Lehrkräfte sehr vielschichtig und differenziert sind. Das bis dato analysierte Datenmaterial wurde erneut auf die neuen Hauptkategorien hin überprüft und entsprechende Textstellen codiert. Durch die Verwendung von MAXQDA konnte die Datenmenge zeiteffizient und systematisch am Computer analyisiert und bearbeitet werden. Das Kategoriensystem (oder Codesystem in MAXQDA) stellt sich wie folgt dar:

Abb. 2: Angewandtes Kategoriensystem.

Codesystem	**2265**
Einschätzung der videographierten Schülerinnen	589
Unterrichtsmethoden	841
Generelle Verfahren zur Diagnose	231
Kenntnis über Diagnoseinstrumente	301
Wünsche für die Weiterbildung	191
Allgemeine Aussagen über Unterricht und Lernende	71
Bildungspolitik/Schulpolitik	41
Sets	**0**

Über die Möglichkeit, in MAXQDA Memos als Gedankenstütze während des Codier-Prozesses zu verfassen, werden zum einen sogenannte Code-Memos und zum anderen sogenannte Dokument-Memos festgehalten. Beide Funktionen dienen als digitales Forschertagebuch, um zum Beispiel Auffälligkeiten von Sprachäußerungen während des Codierens einer Person zuzuordnen, Ankerbeispiele festzuhalten, oder generelle Informationen über die Personen zu dokumentieren.

Im Anschluss an den Prozess des Paraphrasierens können die den Hauptkategorien zugeordneten Codes einer Person in eine Excel-Tabelle exportiert werden, um die weiteren Schritte der *Zusammenfassenden Inhaltsanalyse* zu vollziehen.

Schritt 2-4:
Nachfolgende Abbildung zeigt die systematische Darstellung des Ablaufmodells der *Zusammenfassenden Inhaltsanalyse*, wobei die von Mayring (2008) beschrieben siebenschrittige Analyse auf Grund der Datenfülle verkürzt wurde (vgl. Mayring 2008: 60ff.).

Abb. 3: Beispiel für die Zusammenfassende Inhaltsanalyse.

Hauptkategorie	**Paraphrase**	**Generalisierung**	**Reduktion**	**Sub-Kategorie**	**Textbelegstelle**	**Zeile**
Kenntnis über Diagnoseinstrumente	Peer-Evaluation ist problematisch, da sie sich nicht zuhören	Problem des Zuhörens	Problem: sozialer Faktor	Peer Evaluation	was aber eigentlich so ‚ne Peer Evaluation schon mal deswegen schwierig macht, ist die Tatsache, dass es häufig kaum möglich ist, dass sie einander zuhören	102-102

Im zweiten Schritt des Analyseprozesses werden die Paraphrasen auf ein Abstraktionsniveau weg von umgangssprachlichen Äußerungen der Befragten hin zu verallgemeinernden Aussagen im Sinne einer Generalisierung festgelegt. In Schritt drei erfolgt dann das Streichen und Bündeln von bedeutungsgleichen Codiereinheiten, auch im Sinne einer Reduktion des bis dahin sehr umfangreichen Datenkorpus. Durch das Zusammenstellen aller Textbelegstellen innerhalb eines Interviews mit derselben Bedeutung – auch Text-Retrieval genannt – können dann aus dem Material heraus geordnete und systematisierte Sub-Kategorien gebildet

werden (Schritt 4). Als Unterschied zu anderen Auswertungsmethoden bezieht sich die *Zusammenfassende Inhaltsanalyse* nur im ersten Schritt der Paraphrasierung auf das Originaltranskript und damit die Quellenangaben, konzentriert sich dann jedoch lediglich auf die aus dem Text entnommenen Informationen (zum Beispiel Mayring 2008: 64ff.). Diese Loslösung vom eigentlichen Interviewtranskript und der damit einhergehenden starken Fokussierung auf eigene Rückschlüsse birgt die Gefahr einer frühzeitigen entkoppelten Interpretation, die ohne Rückbezug auf Textbeispiele vollzogen wird. Daher erscheint es mir unumgänglich, die Spalten sechs (Textbelegstelle) und sieben (Transkriptzeile) in die Übersicht mit aufzunehmen, auch um dem Leitprinzip „am Material" weiterhin zu folgen. Insbesondere die siebte Spalte erlaubt es, im Zweifel die Textstelle im Transkript in MAXQDA wiederzufinden und am Material zu kontextualisieren.

Schritt 5:
Nachdem die Sub-Kategorien induktiv im Analyseprozess gebildet worden sind, wird nun das gesamte Datenmaterial in einem weiteren Schritt von einer zweiten Forscherin durch eine ebenfalls induktive Herangehensweise analysiert, um die Qualität des Codierprozesses zu maximieren. Diese Validierung erfolgt nach dem Prinzip des konsensuellen Codierens (vgl. u.a. Schmidt 1997). Um im Anschluss an das unabhängige Codieren einen diskursiven Konsens finden zu können, wird vorab ein Kategoriensystem festgelegt. Hierzu stelle ich meine thematischen Hauptkategorien meiner Forschungspartnerin zur Verfügung, die wiederum in einem individuellen Arbeitsprozess die Transkripte analysiert. Das Verfahren, die eigene Forschung von unterschiedlichen Punkten aus zu betrachten (vgl. Flick 2011: 11) und darüber in einen wissenschaftlich fundierten Dialog zu treten, ist ein wesentliches Qualitätsmerkmal von intersubjektiv nachvollziehbarer Forschung, welche durch eine Forschertriangulation (vgl. u.a. Seipl/Rieker 2003: 225, Flick 2011: 14) gesichert wird. Dadurch können Sub-Kategorien an unklaren Punkten präziser definiert werden. Dies erfolgte zum Beispiel bei der Hauptkategorie „Wünsche für die Weiterbildung", bei der meine Forschungspartnerin und ich weitere Sub-Kategorien festlegten: „Wünsche generell", „Wünsche Material" und „Wünsche Ablauf einer Fortbildung".

Für das konsensuelle Codieren werden zwei unterschiedliche Medien eingesetzt: einerseits werden einige Interviews mittels der *paper-and-pencil* Technik (vgl. Kuckartz 2010: 13) codiert, andererseits wird auf die Software MAXQDA zurückgegriffen. Diese erlaubt es, im Team zu arbeiten, indem verschiedenfarbige Codes entweder Übereinstimmung oder Unterschiede der Forschungspartner signalisieren. Mithilfe dieser farbigen Hervorhebungen kann das Material

zunächst auf Gemeinsamkeiten und Unterschiede hin überprüft werden, um dann in einem zweiten Schritt mit der Forschungspartnerin in einen Dialog zu treten und im Zweifel Sub-Kategorien zu präzisieren.

Schritt 6:
Nachdem die Sub-Kategorien diskutiert und zum Teil präzisiert worden sind, erfolgt im nächsten Schritt eine synoptische und übersichtliche Einordnung des Kategoriensystems (Hauptkategorien und Sub-Kategorien) innerhalb eines Interviews. Des Weiteren wird eine querschnittartige Zusammenstellung des Kategoriensystems aller Interviews vorgenommen, um erste Einsichten in Bezug auf die thematische Vergleichbarkeit der Aussagen und der Sub-Kategorien zu erhalten.

Schritt 7:
Maßgeblich entscheidend für den siebten Schritt ist der Abgleich der deduktiv gebildeten Kategorien (Evaluierungskompetenz, Kompetenzwissen über fremdsprachliches Lesen, Bedingungswissen, technologisches Wissen oder diagnostische Methodenkompetenz, Änderungswissen und Vergleichswissen) mit den zuvor induktiv gebildeten Sub-Kategorien in einer Excel-Tabelle. Dieser Schritt dient zur weiteren Validierung des vorab gebildeten Kategoriensystems, zur Rückbindung der Ergebnisse an die Forschungsfrage sowie als Systematisierung im Hinblick auf das Verfassen des Forschungsberichts.

Schritt 8:
In einem letzten Schritt werden die Haupt- und Sub-Kategorien schriftlich ausgewertet. Hilfreich in diesem Prozess waren folgende vorab generierte Fragen: „Was wird zu welcher Kategorie alles gesagt?“ und „Was wird zu welcher Kategorie nicht gesagt? Was steht bei den befragten Lehrkräften im Vordergrund?“ Diese Leitfragen legen den Schluss nahe, sowohl inter- als auch intratextuell vorzugehen, um dann die Sub-Kategorien entlang der Hauptthemen thematisch zu vergleichen und aufzubereiten.

5. Zusammenschau: *within-method* Triangulation, konsensuelles Codieren und die technischen Möglichkeiten von MAXQDA

Aufgrund der Methoden-Triangulation bei der Datenerhebung war es in der Auswertung möglich, den Datenkorpus dezidiert und vielschichtig zu analysieren. Die Erkenntnisse hinsichtlich der Diagnosekompetenz der Befragten in

Bezug auf das fremdsprachliche Lesen bilden eine große Perspektivenvielfalt ab, welche sowohl das explizit kommunizierte und verfügbare Erfahrungswissen sowie implizit und subjektiv formulierte Schilderungen der Lehrkräfte umfasst. Zum Beispiel gaben die Befragten an, Diagnoseinstrumente (z. B. Portfolio oder Peer Evaluation) trotz persönlichen Dafürhaltens aufgrund zeitlicher oder sozialer Faktoren nicht einzusetzen. Folglich wünschen sie sich Fortbildungen, welche die Praktikabilität solcher Instrumente aufzeigen und sie unmittelbar im Unterricht einsetzbar machen.

Die Auswertung „als Prozeß [sic] der Auseinandersetzung mit dem Material vor dem Hintergrund theoretischer Überlegungen" (Schmidt 1997: 548) zu begreifen, war Ziel einer induktiv-deduktiven Kategorienbildung sowie -auswertung mithilfe der *Qualitativen Inhaltsanalyse*. Da die *Qualitative Inhaltsanalyse* zunächst sehr stark am und mit dem Material arbeitet, gleichzeitig systematisierend den Datenkorpus aufbereitet und reduziert, erlaubt sie „eine methodisch kontrollierte und für jede/n nachvollziehbare Auswertung" (Kuckartz 2012: 174).

Um die Aussagen der Lehrkräfte zusammenzufassen, eignet sich die *Zusammenfassende Inhaltsanalyse* als differenzierte Auswertungsmethode der *Qualitativen Inhaltsanalyse*. Nach der vollständigen Auswertung der Interviews nach Mayring (2008) mittels der am Material gebildeten Kategorien wird dieses Analyseverfahren mit dem der *Inhaltlich Strukturierenden Inhaltsanalyse* verknüpft, um dem Anspruch der intersubjektiven Nachvollziehbarkeit gerecht zu werden und thematische Vergleichbarkeit herzustellen.

Angesichts der Fülle an Datenmaterial erwies sich die Verwendung der Software MAXQDA als äußerst hilfreich. Neben der problemlosen Zuordnung der Paraphrasen zu einzelnen Textabschnitten und der entsprechenden Verortung im Kategoriensystem sind das einfache Hin- und Herwechseln zwischen den einzelnen Interviews, das systematische Verfassen von Memos sowie das Exportieren der Daten in eine Excel-Tabelle mühelos realisierbar. Einzig die Länge der Paraphrasen ist in der Software auf eine bestimmte Zeichenanzahl beschränkt, sodass gelegentlich zwei sogenannte Codes für eine Paraphrase vergeben werden müssen, welche dann in mühsamer Arbeit in Excel wieder zu einer sinnvollen Paraphrase zusammengeführt werden müssen.

Abgesehen von dieser Hürde bietet die Software auch im Hinblick auf das konsensuelle Codieren eine enorme Erleichterung, da es durch diese Art der Validierung möglich ist, die beiden separat entwickelten und induktiv gebildeten Kategoriensysteme aufeinander zu beziehen, um so Gemeinsamkeiten und Unterschiede der unterschiedlichen Forscherperspektiven sichtbar zu machen. Insgesamt stellt die Verwendung der *Qualitativen Inhaltsanalyse*

sowie die Auswertung mit MAXQDA eine geeignete Kombination für Forschungssettings dar, die sich dem Untersuchungsgegenstand zunächst explorativ nähern (induktive Kategorienbildung), um dann mittels einer deduktiven Herangehensweise eine thematische Vergleichbarkeit der einzelnen Interviews herzustellen.

Literatur

BMBF (Bundesministerium für Bildung und Forschung) (Hrsg.) (2007). *Förderung der Lesekompetenz – Expertise*. Bildungsforschung Band 17 Online: http://www.bmbf.de/pub/bildungsreform_band_siebzehn.pdf [04.07.14].

Bogner, Alexander / Menz, Wolfgang (2005). Das theoriegenerierende Experteninterview. Erkenntnisinteresse, Wissensform, Interaktion. In: Bogner, Alexander / Littig, Beate / Menz, Wolfgang (Hrsg.). *Experteninterviews: Theorien, Methoden, Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 33-70.

Bogner, Alexander / Littig, Beate / Menz, Wolfgang (Hrsg.) (2005). *Experteninterviews: Theorien, Methoden, Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Fielding, Nigel G. / Fielding, Jane L. (1986). *Linking Data: The Articulation of Qualitative and Quantitative Methods in Social Research*. London and Beverly Hills: Sage.

Flick, Uwe (2011). *Triangulation. Eine Einführung*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gass, Susan M. / Mackey, Alison (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser, Barney G. / Strauss, Anselm L. (1998). *Grounded Theory: Strategien qualitativer Forschung*. Bern: Huber.

Gläser, Jochen / Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hallet, Wolfgang (2006). *Didaktische Kompetenzen. Lehr- und Lernprozesse erfolgreich gestalten*. Stuttgart: Klett.

Harsch, Claudia / Nold, Günter / Schröder, Konrad (2006). DESI – Die sprachpraktischen Kompetenzen unserer Schülerinnen und Schüler im Bereich Englisch. Zentrale Befunde. *Neusprachliche Mitteilungen aus Wissenschaft und Praxis* 59/3, 11-32.

Hesse, Ingrid / Latzko, Brigitte (2011). *Diagnostik für Lehrkräfte.* 2. Aufl. Opladen: Barbara Budrich.

Hessisches Kultusministerium (kein Datum). *Bildungsstandards und Inhaltsfelder. Das neue Kerncurriculum für Hessen.* Online: http://qualitaetsentwicklung.lsa.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/IQ_Internet/med/4c2/4c22d584-b546-821f-012f-31e2389e4818,22222222-2222-2222-2222-222222222222 [10. 3.2014].

Hochstetter, Johanna (2011). *Diagnostische Kompetenz im Englischunterricht der Grundschule.* Tübingen: Narr.

Ingenkamp, Karl-Heinz (1991). Pädagogischen Diagnostik. In: Roth, Leo (Hrsg): *Pädagogik. Handbuch für Studium und Praxis*. München: Ehrenwirth, 760-785.

KMK (Sekretariat der Ständigen Konferenz der Kultusminister in der Bundesrepublik Deutschland) (Hrsg.) (2004). *Standards für die Lehrerbildung: Bildungswissenschaften.* Online: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [10.3.2014].

Köller, Olaf / Knigge, Micheal / Tesch, Bernd (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.

Kruse, Jan (2011). Reader *Einführung in die Qualitative Interviewforschung.* Unveröffentlichtes Manuskript. Freiburg.

Kuckartz, Udo (2010). *Einführung in die computergestützte Analyse qualitativer Daten*. 3. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kuckartz, Udo (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim und Basel: Juventa.

Kvale, Steinar (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing.* Thousand Oaks: Sage.

Lamnek, Siegfried (2002). Qualitative Interviews. In: König, Eckard / Zedler, Peter (Hrsg.).*Qualitative Forschung*. 2. Aufl. Weinheim, Basel: Beltz, 157-193.

Mayring, Philipp (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 10. Aufl. Weinheim, Basel: Beltz.

McElvany, Nele / Schroeder, Sascha / Hachfeld, Axinja / Baumert, Jürgen / Richter, Tobias / Schnotz, Wolfgang / Horz, Holger / Ullrich, Mark (2009). Diagnostische Fähigkeit von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*. 23/3-4, 223-235.

Meuser, Michael / Nagel, Ulrike (2005). ExpertInneninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In: Bogner, Alexander / Littig, Beate / Menz, Wolfgang (Hrsg.). *Experteninterviews: Theorien, Methoden, Anwendung*. Wiesbaden: Verlag für Sozialwissenschaften, 71-93.

Meuser, Michael / Nagel, Ulrike (2010). Das ExpertInneninterview – Wissensoziologische Voraussetzungen und methodische Durchführung. In: Friebertshäuser, Barbara / Prengel, Annedore (Hrsg.). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim und München: Juventa, 457-471.

Paradies, Liane / Linser, Hans Jürgen / Greving, Johannes (2011). *Diagnostizieren, Fordern und Fördern*. 4. Aufl. Berlin: Cornelsen.

Schmidt, Christiane (1997). ‚Am Material': Auswertungstechniken für Leitfadeninterviews. In: Friebertshäuser, Barbara / Prengel, Annedore (Hrsg.). *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft.* Weinheim und München: Juventa, 544-567.

Schrader, Wolfgang (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie* 23/3-4, 237-245.

Seipl, Christian / Rieker, Peter (2003). *Integrative Sozialforschung. Konzepte und Methode der qualitativen und quantitativen empirischen Forschung*. Weinheim und München: Juventa.

Settinieri, Julia (2014). Forschst Du noch, oder triangulierst Du schon? In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 17-35.

Spinath, Birgit (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie* 19/1-2, 85-95.

Spinner, Kaspar H. (2004). Lesekompetenz in der Schule. In: Artelt, Cordula / Stanat, Petra / Schneider, Wolfgang / Schiefele, Ulrich (Hrsg.): *Struktur, Entwicklung und Förderung der Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS Verlag für Sozialwissenschaften, 127-138.

Steinke, Ines (1999). *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Weinheim, München: Juventa.

Steinke, Ines (2007). Gütekriterien qualitativer Forschung. In: Flick, Uwe / von Kardoff, Ernst / Steinke, Ines (Hrsg.): *Qualitative Forschung. Ein Handbuch*. 5. Aufl. Hamburg: Rowohlt, 319-331.

Südkamp, Anna / Möller, Jens / Pohlmann, Britta (2008). Der Simulierte Klassenraum. Ein Instrument zur Untersuchung von diagnostischer Kompetenz. In: Lankes, Eva-Maria (Hrsg.): *Pädagogische Professionalität als Gegenstand empirischer Forschung*. Münster: Waxmann, 87-97.

Analyse quantitativer und qualitativer Befragungsdaten mit SPSS

Astrid Jurecka

Zur Untersuchung inhaltlicher Forschungsfragen werden in den Erziehungswissenschaften und der Bildungsforschung immer häufiger Befragungsmethoden sowohl aus der quantitativen als auch der qualitativen Forschungstradition herangezogen. Im Rahmen dieses Kapitels sollen Möglichkeiten und Strategien zur gemeinsamen Auswertung und Interpretation solcher Daten aufgezeigt werden. Im ersten Teil wird der Fokus zunächst auf den schriftlichen Fragebogen sowie das qualitative Interview als Beispiele für gängige Befragungsverfahren aus den beiden Forschungstraditionen gelegt. Ferner wird jeweils auf Grundlagen der Testkonstruktion, auf Itemformate, Skalenniveaus sowie das jeweilige Vorgehen bei der Quantifizierung von Daten eingegangen. Der letzte Teil des Beitrags beinhaltet dann eine Darstellung gängiger statistischer Analyseverfahren sowie deren Durchführung mithilfe des Analyseprogramms SPSS.

1. Einleitung

Nachdem etwa seit Mitte der 1920er Jahre in den Sozialwissenschaften zwei methodische Forschungsrichtungen – die qualitative und die quantitative – unterschieden wurden (Kelle 2008), findet sich in den letzten Jahren unter den Begriffen „Methodentriangulation“ und *Mixed Methods* verstärkt wieder die Tendenz, Methoden aus beiden Forschungstraditionen zur Beantwortung und Untersuchung inhaltlicher Fragestellungen heranzuziehen (ebd.). So konstatieren etwa Gläser-Zikuda et al. (2012: 7f.) für die empirische Bildungsforschung und die Erziehungswissenschaften die Etablierung qualitativer Forschungsmethoden, die stärkere Wahrnehmung quantitativer Forschungsmethoden, und die Kombination qualitativer und quantitativer Verfahren als derzeitige forschungsmethodische Trends. Dabei unterscheiden die Autoren bezüglich der Verbindung qualitativer und quantitativer Methoden unterschiedliche Modelle: das Vorstudienmodell, das Verallgemeinerungsmodell, das Vertiefungsmodell sowie das Triangulationsmodell (ebd.: 8). Diese Modelle zeigen unterschiedliche Zeitpunkte und Stadien an, zu denen im Rahmen wissenschaftlicher Studien eine Kombination beider methodischer Richtungen sinnvoll sein kann. So wird etwa im Rahmen des Vorstudienmodells, welches häufig zur Generierung von Hypothesen verwendet wird, eine qualitative der quantitativen Untersuchung vorgeschaltet, und bei Anwendung des Verallgemeinerungsmodells werden die

Ergebnisse einer vollständig durchgeführten qualitativen Studie anhand quantitativer Methoden auf Verallgemeinerung hin überprüft. Im Vertiefungsmodell hingegen werden zunächst quantitative Methoden angewandt, deren Interpretation anschließend unter Verwendung qualitativer Methoden vertieft wird, während im Rahmen des Triangulationsmodells Fragestellungen gleichzeitig anhand qualitativer und quantitativer Methoden bearbeitet werden (ebd.: 8). Ähnliche Vorgehensweisen werden auch bereits im Rahmen der *Grounded Theory* (z.B. Strauss/Corbin 1996) berichtet. Dort wird eine induktive, der Theoriebildung dienende Phase mithilfe offener, qualitativer Fragen einer deduktiven Phase der Theorieüberprüfung und Verallgemeinerung mithilfe von geschlossenen Itemformaten vorgeschaltet (Konrad 2011). Der Vorteil einer Kombination der Forschungsmethoden wird häufig in deren komplementärer Verwendung gesehen (z.B. Flick 2011), um so Schwächen des einen methodischen Ansatzes jeweils gezielt durch Stärken des anderen auszugleichen (Gläser-Zikuda et al. 2012, Kelle 2008, vgl. auch Settinieri in diesem Band).

Wie bereits aus dieser knappen Darstellung kombinierter Anwendungsmöglichkeiten qualitativer und quantitativer Forschungsmethoden ersichtlich wird, ist eine gemeinsame Auswertbarkeit von aus beiden methodischen Traditionen stammenden Daten für alle dargestellten Modelle von großer Relevanz. Das vorliegende Kapitel wird sich daher speziell mit der Aufbereitung und gemeinsamen Analyse von Daten auseinandersetzen, die mithilfe von „qualitativen" oder „quantitativen" Befragungsmethoden erhoben wurden. Dabei wird zur besseren Nachvollziehbarkeit der Inhalte an den entsprechenden Stellen primär auf Standardwerke aus der Methodik verwiesen. Dies soll es dem Leser erleichtern, die hier lediglich übersichtsartig dargestellten Thematiken eigenständig vertiefen zu können.

In der sozialwissenschaftlichen Forschung erfolgt ein großer Teil der Datenerhebung anhand unterschiedlicher Befragungsformen. Im ersten Teil des Beitrags werden zunächst gängige Befragungsformen und Instrumente zur Datenerhebung aus der quantitativen und der qualitativen methodischen Tradition dargestellt. Dabei wird der Fokus vor allem auf dem schriftlichen Fragebogen (als Beispiel für eine traditionell eher quantitative Befragungsmethode) sowie dem Interview (als Beispiel für eine tendenziell eher qualitativ ausgerichtete Befragungsmethode) liegen. Ferner werden die für eine gemeinsame Auswertung der Daten notwendigen Schritte dargestellt. Statistische Analyseverfahren setzen dabei in der Regel voraus, dass die Daten quantifiziert sind, dass ihnen also ein numerischer Wert zugeordnet wurde (vgl. Kempf 2010, Konrad 2011), weshalb auf die Quantifizierung von Daten gesondert eingegangen wird. Im letzten Teil des Kapitels wird als Beispiel für eine computerbasierte Datenanalyse die Analysesoftware SPSS vorgestellt, sowie auf die gängigsten statistischen Analyseverfahren und deren Umsetzung mithilfe des Programms eingegangen.

2. Die Befragung als Methode zur Datenerhebung

Die Befragung ist eine der am häufigsten verwendeten Methoden zur Sammlung von Informationen und Daten in der empirischen Forschung, wobei sich Alltagsfragen von wissenschaftlichen Fragestellungen hinsichtlich ihrer theoretisch-wissenschaftlichen Fundierung und hinsichtlich des Grades der Standardisierung unterscheiden (Konrad 2011: 4). Des Weiteren ist eine Befragung, vor allem bei Verwendung eines Fragebogens, systematisch und zielgerichtet, sie ist gegenstandsbezogen und findet unter kontrollierten Bedingungen statt (ebd.: 7).

Eine Befragung kann entweder mündlich oder schriftlich erfolgen. Im ersten Fall werden dem Befragten durch einen geschulten Interviewer Fragen mündlich gestellt oder vorgelesen, die dieser wiederum in mündlicher Form beantwortet. Dies kann in einer Face-to-Face-Situation erfolgen, oder aber aus der Distanz (vgl. z.B. Broda 2006), beispielsweise über das Telefon oder über Videotelefonie. Im Falle einer schriftlichen Vorgabe liest und bearbeitet der Befragte die vorgegebenen Fragen üblicherweise eigenständig, und auch deren Beantwortung erfolgt in schriftlicher Form. Sowohl mündliche als auch schriftliche Fragen können von offenem oder geschlossenem Format sein. Offenes Format bedeutet, dass dem Befragten keine Antwortmöglichkeiten vorgegeben werden. Dies ist häufig im Rahmen eines Interviews der Fall, es werden jedoch auch offene Antwortformate in Fragebögen verwendet. Geschlossene Antwortformate werden üblicherweise bei der Konstruktion schriftlicher Fragebögen angewandt, hier sind verschiedene konkrete Lösungs- oder Antwortmöglichkeiten jeweils vorgegeben (z.B. Bühner 2006). Eine detailliertere Darstellung unterschiedlicher Item- und Antwortformate erfolgt in den folgenden Abschnitten.

2.1 Der schriftliche Fragebogen

Der schriftliche Fragebogen gehört zu den gängigsten Befragungsmethoden in den Sozialwissenschaften. Obgleich er durchaus qualitativ ausgerichtet sein kann, beispielsweise bei gewähltem offenem Antwortformat, wird der Fragebogen aufgrund seines oft deduktiven Charakters und des meist geschlossenen Antwortformats üblicherweise den quantitativen Methoden der Datenerhebung zugeordnet.

Mithilfe eines Fragebogens können relativ schnell und ökonomisch die Daten einer großen Anzahl von Personen erhoben und ausgewertet werden, weshalb er gut zur Überprüfung der Allgemeingültigkeit von Theorien und Hypothesen geeignet ist. Im Rahmen von Fragebögen werden unter Verwendung geschlossener Antwortformate üblicherweise vorformulierte Antwortmöglichkeiten schriftlich vorgegeben (z.B. Konrad 2011: 51; zu Vor- und Nachteilen verschiedener Itemformate siehe auch Jonkisz et al. 2011: 31ff.), von denen dann die jeweils

am ehesten zutreffende von der befragten Person ausgewählt wird. Fragebögen sind häufig in hohem Maße im Hinblick auf ihre Bearbeitung und Auswertung standardisiert. Nach Bühner (2006: 46) werden bei der Test- und Fragebogenentwicklung unterschiedliche Stufen durchlaufen. Dabei handelt es sich unter anderem um Anforderungsanalyse und Problemstellung, Planung und Literatursuche, Eingrenzung des Merkmals und Arbeitsdefinition, Testentwurf und Überprüfung, Itemanalyse und Itemselektion, Kriterienkontrolle sowie Eichung/ Normierung des Testverfahrens (ebd.: 46ff.).

Der Konstruktion eines Fragebogens geht üblicherweise eine spezifische Problem- bzw. Fragestellung voraus. Denkbar wäre etwa die Forschungsfrage, ob die selbst wahrgenommene muttersprachliche Kompetenz von Kindern mit Deutsch als Zweitsprache auch mit der tatsächlichen Sprachkompetenz in der Muttersprache übereinstimmt. Um diese Fragestellung zu untersuchen, könnte daher (neben einem entsprechenden Sprachtest) ein Fragebogen entwickelt werden, der die wahrgenommene muttersprachliche Kompetenz von Personen erfasst.

Nach Festlegung der Problemstellung sowie Festlegung des zu untersuchenden Merkmals, welches in o.g. Beispiel die wahrgenommene muttersprachliche Kompetenz darstellt, folgen üblicherweise die Eingrenzung des Merkmals sowie eine genaue Definition des zu messenden Konstrukts. Dazu können unterschiedliche Quellen genutzt werden, wie beispielsweise zu der Thematik bereits existierende Theorien, empirische Studien, Testverfahren, oder auch die Durchführung einer Laien- oder Expertenbefragung (z.B. Bühner 2006: 47). Auf diese Weise wird das Merkmal mit allen dazugehörigen Facetten zunächst genau definiert. Darauf basierend werden erste Überlegungen dahingehend angestellt, auf welche Art und Weise das interessierende Merkmal konkret erfasst bzw. erfragt werden könnte. Für eine möglichst vollständige Erfassung des Merkmals „wahrgenommene muttersprachliche Kompetenz" wäre beispielsweise das Abfragen unterschiedlicher Aspekte sprachlicher Teilkompetenzen, wie etwa des Hör- oder Leseverstehens, denkbar.

Anders als bei physikalischen Größen wie dem Gewicht oder der Länge eines Stocks lassen sich viele interessierende Merkmale leider nicht direkt beobachten oder messen. So ist es etwa, bezogen auf das hier verwendete Beispiel, nicht möglich, einem Menschen dessen Einschätzung der eigenen muttersprachlichen Kompetenzen direkt anzusehen. Das interessierende Merkmal bzw. die interessierende Eigenschaft wird daher häufig auch als *latente Variable* bezeichnet. Um also Aussagen bezüglich dieses Merkmals bei einer oder mehreren Personen machen zu können, oder anders ausgedrückt, um das *Konstrukt* zu *operationalisieren* (z.B. Breuer 2010: 40), muss zunächst das Merkmal beobachtbar bzw. sichtbar gemacht werden. Dies geschieht üblicherweise, indem – basierend auf theoretischen Annahmen sowie Ergebnissen bereits existierender empirischer

Studien – das Konstrukt anhand von Fragen, im Folgenden auch bezeichnet als *Items*, möglichst genau abgebildet wird. Die zugrundeliegende Annahme ist hier, dass durch die Beobachtung einer Handlung oder Reaktion einer Person (in diesem Falle die Beantwortung eines Items) Rückschlüsse auf die zugrundeliegende, latente Variable und den Grad der Ausprägung einer Person bezüglich dieses Merkmals gezogen werden können. Zur Messung der wahrgenommenen muttersprachlichen Kompetenzen könnte beispielsweise unter anderem ein Item formuliert werden wie „Wie schätzen Sie ihre muttersprachlichen Fähigkeiten bezüglich des Lesens von literarischen Texten ein?“, zu beantworten auf einer mehrstufigen Antwortskala (z.B. 1 = sehr gering; 5 = sehr gut). Diese beobachtbaren Antworten auf Items werden auch häufig als *manifeste Variable* bezeichnet.

Üblicherweise wird so in einem ersten Entwurf eines Fragebogens ein sogenannter *Itempool* erzeugt, d.h. eine größere Anzahl von Items, von denen – basierend auf den vorherigen Recherchen – angenommen wird, dass sie das interessierende Konstrukt abbilden. Dabei muss darauf geachtet werden, dass alle Facetten eines Konstrukts korrekt repräsentiert sind, da es ansonsten zu einer Konstruktunterrepräsentation, oder aber zur Erhebung konstruktirrelevanter Variablen und somit einer Einschränkung der Validität kommen kann (z.B. Messick 1989).

2.1.1 Itemformate

Im Zuge der Konstruktion eines Items muss ferner festgelegt werden, welches *Format* ein Item haben soll. So existieren, wie bereits erwähnt, offene oder geschlossene Itemformate. Obgleich in schriftlichen Fragebögen auch offene Itemformate verwendet werden können, werden diese tendenziell eher den qualitativen Befragungsformen zugeordnet. Auf offene Antwortformate wird daher im Rahmen der Darstellung qualitativer Befragungsformen eingegangen. Üblicherweise besitzt bei quantitativen Befragungen der größte Teil der Fragen ein geschlossenes Antwortformat (Broda 2006: 70). Das bedeutet, dass die befragte Person sich für eine vorgegebene Antwortalternative entscheiden muss. Beispiele für geschlossene Antworten sind etwa Multiple-Choice- und Richtig-Falsch-Fragen, oder auch Rating-Skalen.

Anhand von Rating-Skalen (auch bezeichnet als Likert-Skalen) soll ein Merkmal hinsichtlich der Stärke seiner Ausprägung erfasst werden. Dazu werden bei der Rating-Skala üblicherweise unterschiedliche Abstufungen bzw. Kategorien benannt. Die befragte Person soll dann wählen, welche Antwortalternative, d.h. welche Merkmalsausprägung, am ehesten auf sie zutrifft. Beispiele für solche Skalen sind die Häufigkeitsskala, die Intensitäts-, die Wahrscheinlichkeits- oder die Bewertungsskala (Bühner 2006: 54).

Bei solchen Itemformaten muss außerdem die Anzahl der Merkmalsabstufungen festgelegt werden. Dies ist dabei prinzipiell dem Testautor überlassen, wobei jedoch beachtet werden sollte, dass nicht eine zu geringe oder eine zu große Anzahl an Antwortkategorien gewählt wird. Üblich ist eine Anzahl von vier oder fünf Antwortkategorien, wobei häufig Skalen ohne eine mittlere Kategorie (beinhaltet oft Formulierungen wie „vielleicht", „manchmal", „mittel", etc.) gewählt werden, um die Ambiguität dieser mittleren Kategorie auszuschließen.

2.1.2 Das Skalenniveau

Ziel und Grundlage des Messens ist die Abbildung des sogenannten empirischen Relativs in ein numerisches Relativ. Nach Bortz und Schuster (2010: 13) lässt sich ein Messvorgang „allgemein dadurch charakterisieren, dass einem Objekt bezüglich der Ausprägung einer Eigenschaft eine Zahl zugeordnet wird". Das zu messende Konstrukt muss mit Beobachtungs- und Messoperationen verknüpft werden (vgl. Breuer 2010) und mithilfe von Testitems, die jeweils wiederum von einem bestimmten Antwortformat sind, operationalisiert werden (z.B. Bühner 2006). „Das Messen ist eine Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe Abbildung eines empirischen Relativs in ein numerisches Relativ ist" (Bortz/Schuster 2010: 16). Diese Abbildung wird auch als *Skala* bezeichnet. Das sogenannte *Skalenniveau* gibt dabei Regeln an, welche Eigenschaften eine bestimmte Skala besitzt, und speziell welche Transformationen von Werten jeweils erlaubt sind, ohne dass Informationen dadurch verloren gehen (Borg/Stauffenbiel 1997). Borg und Stauffenbiel (1997: 6) beschreiben das Skalenniveau als eine auf Hypothesen bezüglich des Zusammenhangs von Werten mit anderen Beobachtungen basierende „Rollenzuweisung".

Das Skalenniveau bezieht sich also auf Messeigenschaften von Items, Skalen und Testkennwerten (Bühner 2006: 73). Die bekanntesten und wichtigsten Skalenniveaus sind die Nominalskala, die Ordinalskala, die Intervallskala und die Verhältnisskala (z.B. Bortz/Schuster 2010). Diese werden im Folgenden kurz dargestellt, da das Skalenniveau für die unten dargestellte Quantifizierung von Daten sowie für die Frage, welche statistischen Auswertungsverfahren jeweils zulässig sind, von Relevanz ist. Dabei schließt jedes höhere Skalenniveau die für niedrigere Niveaus zulässigen Transformationen jeweils mit ein.

Die *Nominalskala* ist das „schwächste" (Borg/Stauffenbiel 1997: 5) Skalenniveau. Üblicherweise werden Antwortformate, bei denen eine Antwort einzelnen wertfreien Kategorien zugeordnet werden kann, dem Nominalskalenniveau zugeordnet (z.B. Bühner 2006: 74). Ein denkbares Beispiel wäre hier etwa die Frage nach der Lieblingsfarbe (Antwortmöglichkeiten: rot, grün, blau, gelb). Hier

können lediglich Aussagen getroffen werden, die angeben, ob die gleiche oder eine andere Kategorie von unterschiedlichen Personen oder Personengruppen gewählt wurde, oder aber Aussagen bezüglich der Häufigkeit des Vorkommens der verschiedenen Antwortkategorien (z.B. die am häufigsten gewählte Farbe ist rot; 20 % haben die Farbe Gelb gewählt, etc.).

Der *Ordinalskala* hingegen werden Daten zugeordnet, anhand derer auch vergleichende Aussagen (besser/schlechter, größer/kleiner, etc.) getroffen werden können (vgl. Bühner 2006). Die Größe der Werte spiegelt auch deren Rangordnung wieder (vgl. Rost 1996: 19), dennoch muss der Abstand zwischen den unterschiedlichen Werten nicht gleich groß sein. Ein denkbares Beispiel wäre hier die Platzierung von Teilnehmern eines 100-Meter-Laufs: Hier wird, je nach Geschwindigkeit, den Personen Platz 1, 2, 3 etc. zugeteilt, allerdings muss der zeitliche Abstand zwischen Person 1 und Person 2 nicht genauso groß sein wie zwischen Person 2 und Person 3.

Intervallskalierte Werte weisen hingegen die gleichen Abstände (Intervalle) auf (z.B. Rost 1996). Ein häufig genanntes Beispiel ist hier die Temperatur in Celsius. Hier sind außerdem Aussagen bezüglich Differenzen möglich, wie z.B. „10 Grad Celsius ist halb so warm wie 20 Grad Celsius", etc. Eine weitere Skala stellt die *Verhältnisskala* dar. Dieser werden viele physikalische Messungen zugewiesen, wie etwa die Länge in cm oder das Gewicht in kg. Hier können Aussagen hinsichtlich der Gleichheit von Verhältnissen getroffen werden (vgl. Bortz/Schuster 2010: 15).

Unterschiedlichen Antwortformaten können unterschiedliche Skalenniveaus zugeordnet werden. Welches Skalenniveau beispielsweise eine Rating-Skala hat, ist dabei bislang nicht abschließend geklärt. In Abhängigkeit davon, ob die Abstände zwischen den Abstufungen der Skala gleich sind oder nicht, kann für die Rating-Skala entweder ein Ordinal- oder Intervallskalenniveau angenommen werden (siehe dazu auch Bühner 2006: 55). Die meisten gängigen statistischen Auswertungsverfahren setzen eine Skalierung mindestens auf Intervallskalenniveau voraus.

2.1.3 Codierung und Quantifizierung

Die Codierung und Quantifizierung von aus geschlossenen Antwortformaten stammenden Daten ist in der Regel relativ problemlos möglich (vgl. Konrad 2011: 89). So liegen häufig Werte wie das Alter oder bestimmte Reaktionszeitwerte bereits in Zahlenform vor. Bezüglich geschlossener Antworten mit mehreren Antwortkategorien werden üblicherweise den Antwortkategorien Zahlen zugeordnet. So kann beispielsweise der Antwortmöglichkeit „stimme nicht zu"

die Zahl eins zugewiesen werden, „stimme eher nicht zu“ die Zahl 2, etc. Wenn die Antwortkategorien dabei eine mehr oder weniger starke Ausprägung eines Merkmals abbilden, dann ist es häufig sinnvoll, der niedrigsten Ausprägung auch die niedrigste Zahl zuzuordnen, da dies die Auswertung und Ergebnisinterpretation vereinfacht. Aber auch bei nominal skalierten und dichotomen Items, wie etwa der Frage nach dem Geschlecht, sollten den Kategorien festgelegte Zahlenwerte zugeordnet werden. So könnte etwa die Kategorie „weiblich“ mit 1, und „männlich“ mit 2 codiert werden. Ist eine Frage einfach mit „Ja“ oder „Nein“ beantwortbar, dann wird „Ja“ üblicherweise die Zahl 1, und „Nein“ die Zahl 0 zugeordnet. Gleiches gilt für Richtig-Falsch-Antworten.

Schwieriger wird die Zuordnung bei Mehrfachantworten. In der Regel sollten sich die unterschiedlichen Werte einer Variablen gegenseitig ausschließen (Janssen/Laatz 2013: 297). Allerdings kommt es vor, dass bei einer Frage mehrere Antwortmöglichkeiten zugelassen sind. In diesem Fall kann eine dichotome Variable pro Antwortmöglichkeit gebildet werden. Jeder dieser Variablen wird dann beispielsweise der Wert 0 oder 1 zugeordnet werden. Auf diese Weise wird codiert, ob eine Antwortmöglichkeit angekreuzt wurde oder nicht (Multiple Dichotomien-Methode; vgl. ebd.). In einigen computerbasierten Analyseprogrammen wie SPSS (SPSS inc./IBM 2012) gibt es außerdem die Möglichkeit, Mehrfachantwortsets zu definieren.

2.2 Das qualitative Interview

Eine andere, häufig verwendete Erhebungsmethode in den Sozialwissenschaften ist das Interview. Dieses wird üblicherweise der qualitativen Forschungsmethodik zugeordnet, weshalb es im Rahmen dieses Beitrags als Beispiel für qualitative Befragungsmethoden ausgewählt wurde. Es existieren dabei unterschiedliche Interviewvarianten, wie etwa das narrative Interview, das ethnografische Interview, das rezeptive Interview, das problemzentrierte Interview, das fokussierte Interview, das episodische Interview, das themenzentrierte Interview, das personenzentrierte Interview, das systemische Interview, das halbstrukturierte Interview oder das Experteninterview (Mey/Mruck 2011: 423ff.). Die Interviewformen unterscheiden sich in ihrem Standardisierungs- und Strukturiertheitsgrad und sind teilweise leitfadenbasiert oder folgen vorformulierten Fragen. Des Weiteren können Interviews anhand der „evozierten Textsorte (Erzählung, Bericht, Argumentation, Sachdarstellung, Kenndaten, Meinungen, etc.) geordnet werden“ (ebd.: 428). Welches Verfahren im Einzelnen ausgewählt wird, ist abhängig von Forschungsinteresse und Anwendungsbereich.

Üblicherweise wird ein Interview von einem Interviewer mit einer zu interviewenden Person durchgeführt. Interviews können synchron (*face-to-face*; telefonisch) oder asynchron (z.B. E-Mail-Interview) durchgeführt werden. Der Interviewer stellt die je nach gewählter Methode zugehörigen Fragen, die Antworten werden meist aufgezeichnet. Dies kann beispielsweise per Audioaufzeichnung, Videoaufzeichnung oder Mitschrift (Notizen/Protokoll) erfolgen (z.B. Konrad 2011). Die so gewonnenen Daten müssen danach entsprechend ausgewertet und ggf. codiert werden. Eine für die Quantifizierung von qualitativen Befragungsdaten gut geeignete Methode ist hier beispielsweise die qualitative Inhaltsanalyse (Mayring 2011: 602): „Sind mit diesen Verfahren Zuordnungen von Kategorien zum Textmaterial regelgeleitet vorgenommen worden, so lassen sie sich gegebenenfalls auch quantitativ […] weiterverarbeiten. Deshalb nimmt die qualitative Inhaltsanalyse im (wenig produktiven) Streit um qualitative oder quantitative Methoden eine gewissen Zwischenstellung ein".

2.2.1 Itemformate

Typisch für qualitative Befragungsformen sind offene Antwortformate. Bei der mündlichen Befragung können, je nach Strukturiertheits- und Standardisiertheitsgrad der gewählten Interviewform, die befragten Personen demnach mehr oder weniger frei auf die durch den Interviewer gestellten Fragen antworten. Denkbare Varianten von Item- bzw. Antwortformaten sind dabei eine vollständig offene Antwort, das Auswählen verschiedener mündlich vorgegebener Antwortmöglichkeiten, oder aber auch diverse Zwischenformen.

2.2.2 Kategorisierung und Codierung qualitativer Interviews

Die Antworten im Rahmen eines mündlichen Interviews werden häufig als Ton- oder Videomitschnitt festgehalten, welcher dann vor der Auswertung transkribiert wird. Die Transkription sollte dabei regelgeleitet erfolgen (vgl. auch Konrad 2011: 47). Ziel der Auswertung ist meist eine Verdichtung des Textes bzw. der darin enthaltenen Aussagen. Dabei kann auch die Auswertung von Interviews mehr oder weniger standardisiert erfolgen (z.B. Schmidt-Atzert/Amelang 2012: 325). Da die Auswertung von offenen Antworten im Rahmen eines Interviews häufig kompliziert und mit größerem Aufwand verbunden ist, werden oft Kategorien festgelegt, anhand derer ein Interview systematisch ausgewertet wird. Eine solche Kategorisierung ermöglicht eine Vergleichbarkeit der Antworten, beispielsweise von unterschiedlichen Gruppen oder zu unterschiedlichen Zeitpunkten (vgl. Konrad 2011: 50). Die Kategorisierung kann sowohl vor als auch nach der Durchführung des Interviews erfolgen, und stellt oft einen separaten Auswertungsschritt dar.

Eine häufig angewandte Auswertungsmethode ist die Inhaltsanalyse, die sich in Verfahren zur Zusammenfassung, Explikation, und Strukturierung unterteilen lässt (z.B. Konrad 2011: 47f.). Beispielsweise bei der unter 2.2 bereits kurz angesprochenen qualitativen Inhaltsanalyse von Mayring (z.B. 2008, 2011) „geht es darum, diese Grundformen des Interpretierens von Text mit inhaltsanalytischen Regeln beschreibbar und überprüfbar werden zu lassen“ (Mayring 2011: 602). Von besonderer Relevanz für die Inhaltsanalyse ist dabei das Kategoriensystem: „Kategorien stellen die Auswertungsaspekte in Kurzform dar […]. Die Kategorien müssen jedoch in der Inhaltsanalyse genau definiert und mit inhaltsanalytischen Regeln muss die Zuordnung zum Text festgelegt werden […]. Dies geschieht in der Regel theoriegeleitet, bei deduktiven Kategorienanwendungen explizit durch die vorab festgelegte Definition der Kategorien, bei induktiver Kategorienbildung durch die Gruppierung der induktiven Kategorien zu Hauptkategorien“ (ebd.: 603). Wird etwa die Vorkommenshäufigkeit solcher Kategorien analysiert, dann kann auch eine quantitative Datenanalyse erfolgen.

Ein relevanter Auswertungsschritt ist dementsprechend die Festlegung von Analyse- und Kodiereinheit (vgl. Konrad 2011: 49). Dabei muss festgelegt werden, welche Textstücke genau auf ihre Einordnung in eine Kategorie, bzw. auf das Vorhandensein und ggf. die Ausprägung des interessierenden Merkmals hin überprüft werden sollen. Konrad (2011: 49) nennt bezüglich der Analyse des Materials drei Aufgaben des/der Forschers/in: die Definition von Kategorien, die Anführung von Ankerbeispielen, und die Festlegung von Kodierregeln. Basierend auf der so erfolgten Auswertung und Codierung des Interviews folgt dann die Interpretation der Ergebnisse. Ein weiterer, dem vorgeschalteter Schritt kann jedoch auch in der statistischen Auswertung und der dazu notwendigen Quantifizierung der Daten liegen.

2.2.3 Quantifizierung qualitativer Daten

Eine Quantifizierung qualitativer Daten bedeutet, dass die „statistische Weiterverarbeitung es erforderlich macht, dass sie in numerische Variablen übersetzt werden“ (Kempf 2010: 1). Zweck einer solchen Quantifizierung ist es häufig, das Gemeinsame über mehrere Interviews hinweg herauszuarbeiten. Zu diesem Zweck erfolgt üblicherweise vor der Quantifizierung die oben bereits erwähnte Kategorisierung zur Verdichtung der Informationen. Die weiteren Auswertungen und die Quantifizierung der Daten erfolgen bei Interviews meist basierend auf einer Transkription. Um das oben verwendete Beispiel der wahrgenommenen muttersprachlichen Kompetenz wieder aufzugreifen, könnte beispielsweise

zunächst einmal von Interesse sein, welche sprachlichen Teilkompetenzen oder welche sprachlichen Handlungen von den Befragten überhaupt wahrgenommen und als wichtig für die allgemeine muttersprachliche Kompetenz erachtet werden. Aufgrund von diesbezüglichen Aussagen können dann Kategorien gebildet werden, wie beispielsweise „Verstehen von Gesprochenem", „Lesen von Tageszeitungen", „Sprechen", etc. Denkbar ist natürlich auch eine Festlegung der Kategorien auf theoretischer Basis bereits vor der Durchführung des Interviews.

Wurden die Kategorien festgelegt, kann nun zunächst einfach gezählt werden, wie häufig eine bestimmte Kategorie insgesamt gewählt wurde. Hier besteht die Quantifizierung der Daten also in der Erhebung der Häufigkeiten. Dies ist bereits eine wichtige Information, kann sie doch dem/der Forscher/in einen Eindruck vermitteln, ob die gewählten Kategorien für die Befragten tatsächlich von Belang sind und welche der Kategorien bzw. welche der sprachlichen Teilkompetenzen die größte oder geringste Relevanz aufweisen. Ferner können, basierend auf diesen Ergebnissen, die Kategorien auch überarbeitet, ergänzt oder adaptiert und dann erneut ausgezählt werden.

Diese relativ basalen Informationen können bereits von inhaltlicher Relevanz sein. Bezogen auf das Sprachkompetenz-Beispiel wären möglicherweise didaktische Implikationen für die Gestaltung des muttersprachlichen Unterrichts denkbar, etwa wenn durch die Häufigkeit der Nennungen der einzelnen wahrgenommenen Sprachkompetenzen ersichtlich wird, dass eigentlich sprachlich hoch relevante Teilkompetenzen von den Befragten gar nicht als relevant wahrgenommen werden und deren expliziterer Einbezug in den Unterricht daher möglicherweise sinnvoll wäre.

Neben der Erhebung von einfachen Häufigkeiten ist es, je nach Merkmal, auch möglich, die gewählten Kategorien weiter zu unterteilen. Oft ist dies sinnvoll, wenn neben dem bloßen Vorkommen außerdem der Ausprägungsgrad eines Merkmals interessiert. Bezogen auf das Beispiel wäre hier eine Abstufung des interessierenden Merkmals, etwa der sprachlichen Teilkompetenz „wahrgenommenes Hörverstehen", in unterschiedliche Kompetenzstufen[1] denkbar. In diesem Fall würden für eine Quantifizierung der Daten den unterschiedlichen Kompetenzstufen einer sprachlichen Teilkompetenz unterschiedliche Zahlen zugeordnet, beispielsweise der niedrigsten Stufe die Zahl 1 und der höchsten die

---

1 Hier sind dabei nicht empirisch gebildete Kompetenzstufen im Rahmen einer Kompetenzmodellierung gemeint, sondern eine Einordnung von Antworten in theoretisch angenommene Fähigkeitsniveaus.

Zahl 6. Auf diese Art und Weise kann die im Interview enthaltene Information verdichtet, quantifiziert und statistisch ausgewertet werden.

2.2.4 Das Skalenniveau

Sind qualitative Daten quantifiziert, können prinzipiell, je nach Skalenniveau, die gleichen Auswertungsmethoden herangezogen werden wie bei Daten, die anhand quantitativer Methoden gewonnen wurden. Welche statistischen Auswertungsmethoden genau gewählt werden können und sollten, hängt zum einen mit der Art der Forschungsfrage und der damit einhergehenden Hypothese zusammen (z.B. Zusammenhangshypothese oder Unterschiedshypothese). Zum anderen legt das Skalenniveau der Daten fest, welche Verfahren überhaupt zulässig sind. So setzen viele Verfahren der deskriptiven Statistik sowie der Inferenzstatistik mindestens Intervallskalenniveau voraus. Sind beispielsweise bei der Kategorisierung mehrere Abstufungen gewählt worden, anhand derer eine mehr oder weniger starke Ausprägung des gemessenen Merkmals codiert wird, kann hier, genauso wie bei den anhand quantitativer Methoden erhobenen Daten, zumindest von einem Ordinalskalenniveau ausgegangen werden. Im Prinzip ist auch eine Einordnung auf Intervallskalenniveau möglich, nämlich dann, wenn theoretisch davon ausgegangen wird, dass die einzelnen Abstufungen des Merkmals die gleichen Abstände besitzen. Auf die Frage, welche Analyseverfahren für welche Skalenniveaus zulässig sind, wird im folgenden Teil des Kapitels näher eingegangen.

3. Computerbasierte Datenanalyse am Beispiel von SPSS

In diesem Teil des Kapitels soll am Beispiel des Analyseprogramms SPSS (IBM SPSS Statistics, derzeit Version 22; SPSS inc./IBM 2014) auf die computerbasierte Auswertung empirischer Daten eingegangen werden. Prinzipiell sind alle beschriebenen Analysen jedoch auch mit anderen Programmen wie beispielsweise der Open-Source-Software R (http://www.r-project.org) durchführbar. Bei SPSS handelt es sich um eine der gängigsten Statistik- und Datenanalyse-Software (z.B. Janssen/Laatz 2013). SPSS ist ein proprietäres Programm und prinzipiell für unterschiedliche Betriebssysteme wie Windows, Linux oder MacOS erhältlich. Das Programm ist modular aufgebaut; zusätzlich zur Basissoftware, welche die gängigsten statistischen Auswertungsverfahren enthält, existieren spezielle Softwaremodule, die jedoch für das vorliegende Kapitel nicht von Relevanz sind. In SPSS können die Befehle entweder über Dropdown-Menüs oder aber per Syntax eingegeben werden. Im vorliegenden Kapitel wird ausschließlich auf die Verwendung der Menüs eingegangen, zur Verwendung von Syntax wird an

dieser Stelle auf entsprechende Fachliteratur (z.B. Brosius 2013, Janssen/Laatz 2013) verwiesen.

3.1 Grundlagen

Die Verwendung von SPSS setzt für den größten Teil der Analysemethoden eine Quantifizierung aller Daten voraus. Das bedeutet, um SPSS für eine gemeinsame Auswertung der anhand von qualitativen und quantitativen Erhebungsmethoden gewonnenen Daten nutzen zu können, müssen – wie oben beschrieben – allen Merkmalen, Merkmalsausprägungen und Kategorien Zahlen zugeordnet werden. Sind die Daten einmal erhoben und quantifiziert, müssen sie zunächst in die SPSS-Datenmatrix eingegeben und gegebenenfalls für die Analyse vorbereitet werden. Das Programm verfügt dazu über zwei unterschiedliche Ansichten: die *Variablenansicht* und die *Datenansicht* (Abb. 1). Üblicherweise werden bei der Erstellung der Datenmatrix zunächst in der Variablenansicht alle Variablen definiert sowie deren Eigenschaften festgelegt. Hier können unter anderem der Name der Variablen, der Variablentyp (beispielsweise numerisch oder String), das Variablenlabel, sowie Wertelabels (d.h. welche Ausprägung einer Variablen wurde mit welcher Zahl codiert?) festgelegt werden. Pro Frage (Fragebogen) bzw. Kategorie (Interview) wird eine Variable definiert. Ist diese Definition der Datenmaske erfolgt, werden im nächsten Schritt die Daten in der Datenansicht eingegeben. Dort stellt jede Spalte eine der zuvor definierten Variablen dar, und jede Zeile einen Fall, das heißt beispielsweise eine befragte Person. In den einzelnen Zellen sind somit die Probandenrohwerte abgetragen, d.h. jeweils die Antwort einer einzelnen Person auf ein einzelnes Item.

Ist die Dateneingabe erfolgt, ist es oft notwendig, vor der Durchführung der geplanten Analysen die Daten weiter aufzubereiten. Das beinhaltet häufig die *Umcodierung* von bereits eingegebenen Daten. Eine solche Umcodierung könnte etwa dann notwendig sein, wenn eine Frage im Fragebogen nicht so formuliert wurde, dass eine stärkere Zustimmung zu einer Aussage auch tatsächlich mit einer höheren Ausprägung des Merkmals einhergeht, d.h. wenn zur Erfassung eines Merkmals sowohl positiv als auch negativ formulierte Items eingesetzt wurden. Ein weiterer Grund für eine Umcodierung kann die Entscheidung sein, aus theoretischen Gründen bestimmte Variablenausprägungen zusammenzufassen und diesen den gleichen Zahlenwert zuzuordnen (Beispiel: Festlegung bestimmter Altersgruppen, Zusammenfassen von Berufsgruppen, Sprachgruppen etc.). Eine solche Umcodierung kann in SPSS jederzeit relativ unproblematisch über das Menü *Transformieren* und eines der Untermenüs *Umcodieren* erfolgen.

Abb. 1: Datenstruktur in SPSS (Datenansicht; Screenshot); Daten entnommen aus dem SPSS Beispieldatensatz „GSS93“.

GSS93 subset.sav [DatenSet2] - IBM SPSS Statistics Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

Sichtbar: 97 von 97 Variablen

	id	wrkstat	marital	agewed	sibs	childs	age	birthmo	zodiac	educ	degree	padeg	madeg	sex
1	1	1	3	20	3	1	43	5	2	11	1	7	0	1
2	2	1	5	0	2	0	44	8	6	16	3	0	0	1
3	6	5	1	25	2	2	83	3	12	11	1	0	0	1
4	8	5	1	24	3	2	75	11	9	12	1	0	0	1
5	9	1	3	22	1	2	31	7	4	18	4	1	1	1
6	18	1	1	22	9	0	55	1	10	15	1	1	0	1
7	20	1	1	24	2	2	34	4	2	16	3	0	1	1
8	23	5	2	25	2	2	80	5	2	18	4	7	0	1
9	24	1	5	0	5	1	32	7	5	16	3	3	3	1
10	25	2	1	32	7	4	37	9	6	5	0	0	0	1
11	31	1	5	0	1	0	29	5	2	16	4	3	3	1
12	35	5	1	21	3	4	66	2	12	12	1	0	0	1
13	39	2	1	28	7	3	41	9	6	6	0	0	0	1
14	41	1	1	24	4	2	34	1	10	12	1	0	8	1
15	45	1	5	0	2	0	31	3	1	14	1	0	0	1
16	47	1	1	26	1	1	28	4	1	18	4	0	0	1
17	51	1	5	0	1	0	48	6	3	16	3	8	1	1
18	53	1	3	17	21	8	39	3	12	19	4	3	4	1
19	54	8	5	0	3	0	20	6	4	13	1	1	0	1
20	56	1	5	0	1	0	36	10	7	12	1	7	0	1
21	57	5	1	24	6	2	78	2	11	15	3	0	0	1
22	61	1	1	24	5	3	39	4	2	12	1	0	0	1
23	66	1	1	22	3	2	45	11	9	14	1	1	0	1
24	67	1	1	23	1	3	46	8	6	20	4	0	7	1
25	70	1	5	0	3	0	27	6	4	13	1	1	1	1
26	72	1	1	22	2	2	31	5	3	16	3	3	2	1
27	74	5	1	22	3	3	77	8	5	16	3	0	0	1
28	75	2	1	23	2	3	41	11	8	13	1	0	8	1
29	77	5	2	25	5	1	81	11	8	8	0	0	8	1
30	78	4	1	25	8	3	41	12	9	12	1	1	1	1
31	85	1	1	22	0	0	35	7	4	16	3	1	1	1
32	86	1	1	25	3	1	36	7	4	14	1	0	1	1
33	87	1	1	27	2	2	41	12	10	12	1	1	1	1
34	89	1	1	21	1	2	44	11	9	12	1	0	1	1
35	91	1	1	30	3	9	49	2	11	15	1	8	0	1
36	96	1	1	26	1	2	45	6	3	18	4	4	1	1

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit

Eine weitere, häufig notwendige Maßnahme vor weiteren Datenanalysen ist das *Berechnen* von neuen Variablen. Dies kann etwa dann notwendig werden, wenn aus unterschiedlichen Items, die das gleiche Konstrukt erfassen sollen (beispielsweise die muttersprachliche Lesekompetenz), ein Gesamttestwert gebildet werden soll. Dies kann über den Menüpunkt *Transformieren* und das Untermenü *Variable berechnen* erfolgen. Dort können aus bestehenden Variablen neue errechnet werden, wie beispielsweise ein Gesamttestwert mittels einer einfachen Addition derjenigen Variablen, die zu diesem Wert beitragen sollen. Die neue Variable erscheint nach Abschluss der Berechnung in der Datenmatrix und kann für weitere Analysen verwendet werden.

In SPSS erfolgen die Angaben sämtlicher durchgeführter Aktionen sowie die Darstellung der Analyseergebnisse im sogenannten *Ausgabefenster.* Dieses wird automatisch vom Programm erstellt und geöffnet, etwa wenn eine Datei zum ersten Mal in der Sitzung geöffnet oder eine neue Datei gespeichert wird. Der Inhalt des Ausgabefensters muss separat gespeichert werden und kann in unterschiedliche Formate exportiert werden. Ist die Datenaufbereitung abgeschlossen, können die

ersten Auswertungen erfolgen. Welche Verfahren jeweils anzuwenden sind, ist abhängig von Fragestellung und Hypothese (z.B. Unterschiedshypothese, Zusammenhangshypothese), dem Skalenniveau der Daten, sowie der Stichprobengröße (n).

3.2 Datenauswertung

3.2.1 Erste Auswertungsschritte

In einem ersten Schritt werden üblicherweise die Häufigkeitsverteilungen sowie die deskriptiven Statistiken der interessierenden Variablen betrachtet. Hierzu zählen Maße der zentralen Tendenz (Mittelwert/arithmetisches Mittel, Median, Modalwert) und Dispersionsmaße (Varianz, Standardabweichung, Minimum, Maximum, Range).

Häufigkeiten bzw. Häufigkeitsverteilungen sowie unterschiedliche dazugehörige Diagramme können in SPSS über den Menüpunkt *Analysieren* und dort über die Untermenüs *Deskriptive Statistiken* und *Häufigkeiten* angefordert werden. Am Ergebnis ist beispielsweise abzulesen, wie häufig welche Antwort bezüglich eines Items gewählt wurde, oder wie häufig eine bestimmte Gruppe in der Stichprobe vertreten ist (z.B. Männer, Frauen; siehe Tabelle 1).

Die Analyse der deskriptiven Statistiken erfolgt über die Menüpunkte *Analysieren – Deskriptive Statistiken – deskriptive Statistik* (Tabelle 2), oder aber über *Analysieren – Deskriptive Statistiken – Explorative Datenanalyse*. Des Weiteren existiert auch im Rahmen der meisten inferenzstatistischen Auswertungsverfahren in SPSS die Möglichkeit, diese Kennwerte mit ausgeben zu lassen. Mithilfe der deskriptiven Statistiken ist es möglich, erste Kenntnisse über die Verteilung der Daten zu erlangen. So lässt sich eine Aussage darüber machen, ob die Verteilung eher breit- oder schmalgipflig ist (Kurtosis), ob sie beispielsweise linksschief und rechtssteil ist (Schiefe) oder symmetrisch. Die Verteilungsform ist von großer Relevanz, da sie Einfluss auf weitere Auswertungen haben kann. Ist die Varianz beispielsweise sehr klein, da die getesteten Personen sich im Hinblick auf das gemessene Merkmal kaum unterscheiden, dann kann dies u.a. zu einer niedrigeren Korrelation mit anderen Variablen führen (Varianzeinschränkung; z.B. Schmidt-Atzert/Amelang 2012: 10).

Tabelle 1: Häufigkeiten der Variablen „Geschlecht“ (Ausprägungen: männlich/weiblich)

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	37	49,333	49,333	49,333
	weiblich	38	50,667	50,667	100
	Gesamt	75	100	100	100

Tabelle 2: Deskriptive Statistiken der Variablen „muttersprachliche Kompetenz" und „wahrgenommene muttersprachliche Kompetenz"

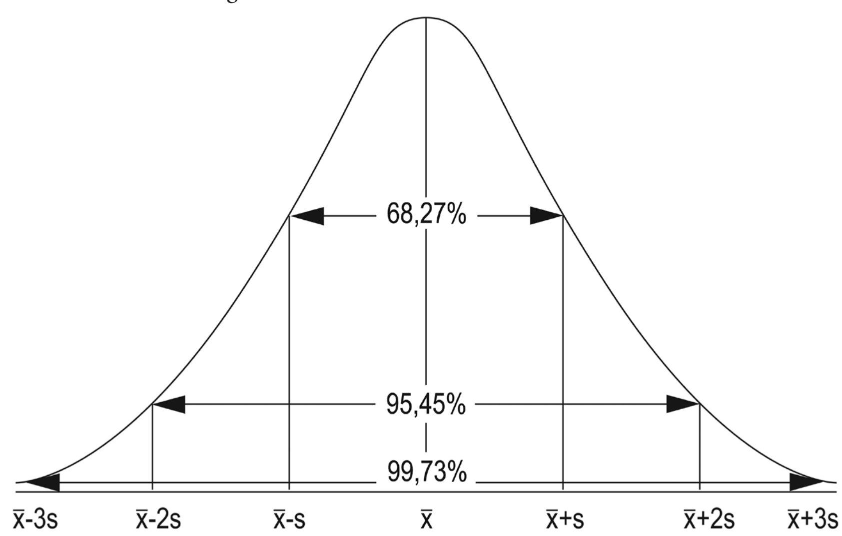
	N	Minimum	Maximum	Mittelwert	Standard-abweichung
Muttersprachliche Kompetenz [1]	75	15	49	33,50	5,3
Wahrgenommene muttersprachliche Kompetenz [2]	75	1	5	3,01	0,9

[1] Max. Wert = 50 Punkte
[2] Angenommene Intervallskalierung des Merkmals

Wie bereits erwähnt, setzen außerdem viele der gängigen Analyseverfahren eine *Normalverteilun*g (NV) der Daten voraus. Die Normalverteilung (Abb. 2) zeichnet sich dadurch aus, dass sich die Daten symmetrisch und glockenförmig verteilen. Des Weiteren sind in dem Bereich +/- 1 Standardabweichung vom Mittelwert ca. 68% der Messwerte zu finden, und +/-2 Standardabweichungen vom Mittelwert ca. 95,5% (z.B. Rasch et al. 2006: 32).

In SPSS können Daten daraufhin überprüft werden, ob sie hinsichtlich ihrer Verteilung statistisch signifikant von einer Normalverteilung abweichen. Dies erfolgt mithilfe eines sogenannten statistischen Tests. Für statistische Analysen gilt allgemein Folgendes:

Abb. 2: Normalverteilung.



Ziel in der „Inferenzstatistik" (oder auch „schließenden Statistik") ist das Ziehen von Rückschlüssen von Stichprobendaten auf eine entsprechende (unbekannte, da nicht komplett erhobene) Population. Bezogen auf unser Beispiel könnte etwa von Interesse sein, ob sich Männer und Frauen hinsichtlich der Wahrnehmung ihrer muttersprachlichen Kompetenz unterscheiden (Unterschiedshypothese). Ferner wird angenommen, dass die Frauen ihre Kompetenz signifikant höher einschätzen als es bei den Männern der Fall ist. Es handelt sich hier also um eine gerichtete Hypothese, da Annahmen bezüglich der Richtung des Unterschieds getätigt werden. Existieren solche Annahmen hingegen nicht, handelt es sich um eine ungerichtete Hypothese.

Als nächstes wird diese wissenschaftliche Hypothese in eine sogenannte statistische Hypothese überführt (Bortz/Schuster 2010: 98). Die Nullhypothese (H0) postuliert dabei, dass in der Population kein Unterschied zwischen den beiden Mittelwerten der beiden Gruppen (genauer: zwischen den Populationsmittelwerten, bezeichnet als $\mu1$ und $\mu2$) existiert. Dies wird formal ausgedrückt als $\mu1= \mu2$. Die sogenannte Alternativhypothese (H1) hingegen drückt aus, dass ein Unterschied zwischen den beiden Populationsmittelwerten existiert (gerichtet: $\mu1< \mu2$ oder $\mu1>\mu2$; ungerichtet: $\mu1\neq\mu2$). Es wird nun –möglichst basierend auf einer repräsentativen Stichproben– mithilfe statistischer Tests überprüft, welche der beiden Hypothesen bestätigt bzw. zurückgewiesen werden soll. Dies geschieht, indem „die Vereinbarkeit der Nullhypothese mit der Empirie" überprüft wird (ebd.: 99).

Die Überprüfung statistischer Hypothesen erfolgt mithilfe der Berechnung einer sogenannten Prüfgröße (z.B. p, F, T je nach angenommener Verteilung der Daten; gewonnen aus den empirisch erhobenen Stichprobendaten), die mit einem sog. „kritischen Wert" abgeglichen werden. Der jeweils zu verwendende kritische Wert kann anhand spezieller Tabellen (z.B. in Rasch et al. 2006: 194ff.) abgelesen werden. Üblicherweise gilt: Überschreitet die Prüfgröße den kritischen Wert, ist von einer statistischen Signifikanz auszugehen, die Nullhypothese (H0) wird verworfen und die Alternativhypothese (H1) angenommen. Der in SPSS mit „Signifikanz" bezeichnete Wert gibt die Wahrscheinlichkeit an, mit der ein Fehler begangen wird, wenn die Nullhypothese verworfen wird. Konventionell wird bei einem Signifikanz-Wert kleiner als 0,05 von einem statistisch relevanten Ergebnis ausgegangen, da so davon ausgegangen werden kann, dass in maximal 5% der Fälle ein Fehler gemacht wird, wenn die H0 verworfen und die H1 angenommen wird.[2]

2 Im Rahmen des Hypothesentestens wird von zwei möglichen Fehlerarten ausgegangen: „Ein Fehler 1. Art wird begangen, wenn eine richtige Nullhypothese zugunsten

Die Überprüfung von Daten auf Normalverteilung erfolgt nun in SPSS über das Menü *Analysieren -Deskriptive Statistiken - Explorative Datenanalyse.* In dem dann geöffneten Auswahlfeld muss nach der Auswahl der zu analysierenden Variablen der Button *Diagramme* angeklickt, und dann ein Häkchen bei *Normalverteilungsdiagramm mit Test* gesetzt werden. Die Ergebnisausgabe beinhaltet dann neben den deskriptiven Statistiken auch Tests auf Normalverteilung: den Kolmogorov-Smirnoff-Test, sowie den Shapiro-Wilks-Test bei n<50. Für die schnelle Interpretation ist hier vor allem der Wert in der Tabellenspalte *Signifikanz* von Interesse: Ist dieser kleiner als 0,05, dann weicht die Verteilung signifikant von der Normalverteilung ab. Anders ausgedrückt: Die Nullhypothese, dass die Daten nicht von einer Normalverteilung abweichen, wird zurückgewiesen und somit die Alternativhypothese H1 angenommen. Bezüglich der Interpretation der übrigen von SPSS ausgegebenen Werte wird aus Platzgründen auf entsprechende Standardwerke (z.B. Janssen/Laatz 2013) verwiesen. Sind die Daten nicht normalverteilt, können teilweise sogenannte nichtparametrische Tests zum Einsatz kommen, auf die an dieser Stelle jedoch nicht näher eingegangen werden soll.

3.2.2 Überprüfung von Zusammenhangshypothesen

Häufig beinhalten wissenschaftliche Fragestellungen die Frage nach dem Zusammenhang zwischen zwei Variablen. Eine auf unser Beispiel bezogene Fragestellung wäre etwa „Existiert ein (statistisch signifikanter) Zusammenhang zwischen der wahrgenommenen und der (per Test erfassten) tatsächlichen muttersprachlichen Kompetenz?“. Eines der zur Bearbeitung solcher Fragestellungen am häufigsten eingesetzten Analyseverfahren ist die *Korrelation*. Hier ist zunächst zu unterscheiden zwischen der bivariaten und der multiplen Korrelation: Erstere betrachtet den Zusammenhang zwischen zwei (intervallskalierten) Variablen, letztere den simultanen Zusammenhang von einer Variablen mit mehreren anderen. Des Weiteren existieren, je nach Skalenniveau der in die Analyse einbezogenen Daten, unterschiedliche Korrelationskoeffizienten. Sind die Daten mindestens auf Intervallskalenniveau und in der Grundgesamtheit normalverteilt, dann wird üblicherweise der *Pearson Produkt-Moment-Korrelationskoeffizient* (r) verwendet. Sind diese Voraussetzungen nicht gegeben, kann ein *Rangkorrelationskoeffizient* (r_s) verwendet

der Alternativhypothese abgelehnt wird. Ein Fehler 2. Art wird begangen, wenn eine falsche Nullhypothese beibehalten wird“ (Bortz/Schuster 2010: 100). Die Fehler werden auch als alpha- und beta-Fehler bezeichnet.

werden. Bei SPSS stehen hier für ordinalskalierte Variablen der Spearman-Koeffizient (r_s) sowie Kendall's Tau-b (T_b) zur Verfügung (z.B. Brosius 2013).

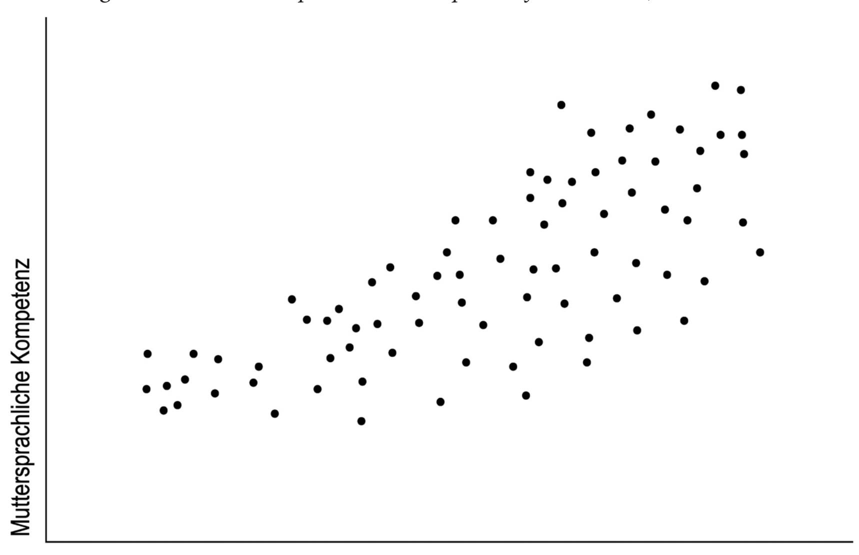
Korrelationskoeffizienten können generell Werte zwischen 1 und −1 annehmen. Dabei gilt: Je näher sich der Wert an Null befindet, desto geringer ist der (lineare) Zusammenhang (z.B. Bortz/Schuster 2010: 125). Dementsprechend bedeutet ein Wert von Null, dass kein linearer Zusammenhang zwischen den Variablen besteht. Je näher der Wert hingegen an 1 oder -1 liegt, desto größer ist der lineare Zusammenhang zwischen zwei Variablen ($r = 1$ bzw. $r = -1$: perfekter linearer Zusammenhang).

Ein positiver Korrelationskoeffizient lässt sich dabei inhaltlich so interpretieren, dass größere Werte auf der einen Variablen systematisch mit größeren Werten auf der anderen Variablen einhergehen („je mehr desto mehr"; zur Herleitung und Berechnung des Korrelationskoeffizienten siehe z.B. Rasch et al. 2006: 125ff.). Bezogen auf unser Beispiel könnte ein positiver Koeffizient von $r = 0,7$ etwa bedeuten, dass eine hohe wahrgenommene muttersprachliche Kompetenz relativ systematisch mit einer hohen getesteten muttersprachlichen Kompetenz einhergeht. Ein negativer Koeffizient bedeutet hingegen, dass ein hoher Wert auf der einen Variablen systematisch mit einem niedrigen Wert auf der anderen Variablen einhergeht („je mehr desto weniger"). Bei der inhaltlichen Interpretation von Korrelationen ist zu beachten, dass diese nicht kausal interpretiert werden dürfen, da keine Aussage über die Richtung eines Zusammenhangs gemacht werden kann. Es wird lediglich ausgesagt, dass ein mehr oder weniger starker linearer Zusammenhang zwischen zwei Variablen existiert bzw. nicht existiert. Bezüglich der Interpretation gibt z.B. Brosius (1998) als Orientierungshilfe an, dass Koeffizienten zwischen $r = 0$ und $r = 0,2$ als sehr schwache Korrelation, $r = 0,2$ und $r = 0,4$ als schwache Korrelation, zwischen $r = 0,4$ und $r = 0,6$ als mittlere, und darüber hinaus als starke bzw. sehr starke Korrelation zu interpretieren sind.

Ein erster Eindruck bezüglich der Stärke des linearen Zusammenhangs kann außerdem über die Erstellung eines Streudiagramms (Abb. 3) gewonnen werden. Dort werden für jede Person bzw. für jeden Fall anhand eines Punktes die Werte der beiden erfassten Variablen auf der x- und y- Achse eines Koordinatensystems abgetragen. Ein perfekter Zusammenhang würde sich nun dadurch zeigen, dass alle Punkte perfekt auf einer 90°-Geraden liegen, deren Richtung wiederum anzeigt, ob es sich dabei um einen positiven, oder um einen negativen Zusammenhang handelt (Gerade von unten links nach oben rechts: positiver Zusammenhang; von oben links nach unten rechts: negativer Zusammenhang). Je näher die Punkte nun um diese Gerade streuen, desto größer ist der lineare Zusammenhang. Ein solches Diagramm kann in SPSS beispielsweise über

den Menüpunkt *Diagramme –veraltete Dialogfelder – Streu-/Punkt-Diagramm* erstellt werden.

Abb. 3: Beispiel Streudiagramm, Darstellung des Zusammenhangs zwischen tatsächlicher und wahrgenommener muttersprachlicher Kompetenz (fiktive Daten).



Beispiel: Positiver Zusammenhang zwischen wahrgenommenen und tatsächlichen muttersprachlichen Kompetenzen (fiktives Beispiel)

Die Berechnung der Korrelationskoeffizienten an sich erfolgt in SPSS über die Menüpunkte *Analysieren – Korrelation – bivariate Korrelation*. In dem dann geöffneten Dialogfeld können die interessierenden Variablen sowie die zu verwendenden Korrelationskoeffizienten gewählt werden. Prinzipiell ist es möglich, gleichzeitig die Korrelationen zwischen mehreren Variablen berechnen zu lassen. Es sollte jedoch bedacht werden, dass bei einer größeren Anzahl von Variablen die im Ausgabefenster erstellte Korrelationstabelle schnell unübersichtlich wird.

Im Ausgabefenster wird dann eine Tabelle (Korrelationstabelle, siehe Tabelle 3) erzeugt, welche die bivariate Korrelation jeder Variablen mit jeder anderen gewählten Variablen anzeigt. In den daraus resultierenden Zellen wird jeweils die Größe des Korrelationskoeffizienten als Maß für die Stärke des Zusammenhangs zweier Variablen angegeben, der Signifikanz-Wert p, sowie die Anzahl der zur

Berechnung verwendeten Fälle. Auch hier gibt der Signifikanz-Wert Auskunft darüber, ob ein Zusammenhang statistisch signifikant ist: Ist dieser Wert kleiner als 0,05 ($p < 0,05$), dann wird die Null-Hypothese „Es besteht kein (statistisch signifikanter) Zusammenhang“ zurückgewiesen und die H1 „Es besteht ein (statistisch signifikanter) Zusammenhang“ angenommen.

Tabelle 3: Korrelationen zwischen den Variablen „muttersprachliche Kompetenz“ und „wahrgenommene muttersprachliche Kompetenz“

		Muttersprachliche Kompetenz	Wahrgenommene Muttersprachliche Kompetenz
Muttersprachliche Kompetenz	Korrelation nach Pearson	1	,7*
	Signifikanz (2-seitig)		,023
	N	75	75
Wahrgenommene Muttersprachliche Kompetenz	Korrelation nach Pearson	,7*	1
	Signifikanz (2-seitig)	,023	
	N	75	75

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.
N = Anzahl der Personen

Neben der bivariaten Korrelation existieren noch weitere Möglichkeiten, Zusammenhänge zwischen Daten zu untersuchen. Möchte man beispielsweise überprüfen, ob es einen Zusammenhang zwischen zwei nominalskalierten Variablen (oder auch zwischen einer nominal- und einer ordinalskalierten Variablen) gibt, so kann dies mittels einer *Kreuztabelle* und dem *Chi²-Test* geschehen. Dabei wird abgeglichen, ob die *beobachteten Häufigkeiten* in den Zellen von den unter Unabhängigkeit der Daten (d.h. unter der Annahme, dass kein Zusammenhang existiert) *erwarteten Häufigkeiten* signifikant abweichen. Für eine detailliertere Darstellung und Interpretation der Kreuztabelle wird auf Janssen/Laatz (2013) oder Brosius (2013) verwiesen.

Die Erstellung der Kreuztabelle erfolgt in SPSS über die Menüpunkte *deskriptive Statistiken – Kreuztabellen*. Im folgenden Dialogfenster kann nun unter anderem festgelegt werden, welche Variablen in den Zeilen und welche in den Spalten der Tabelle erscheinen sollen. Des Weiteren kann unter Verwendung des Buttons „Statistiken“ gewählt werden, welche Zusammenhangskoeffizienten (je nach Skalierung der Variablen) mit ausgegeben werden sollen. Neben dem Chi²-Test und dem Korrelationskoeffizienten handelt es sich dabei beispielsweise um den Kontingenzkoeffizienten für nominalskalierte Variablen oder den bereits erwähnten Kendall’s Tau-b-Koeffizienten für ordinalskalierte Variablen. Der Chi²-Test gibt zwar an, ob ein Zusammenhang

signifikant ist, jedoch sagt er nichts über dessen Stärke aus. Daher sollte einer der genannten Koeffizienten mit hinzugezogen werden. Der in der Ergebnistabelle mit ausgegebene Signifikanz-Wert der Koeffizienten ist wie üblich zu interpretieren.

Ein weiteres gängiges Analyseverfahren zum Betrachten von Zusammenhängen ist die (*multiple*) *lineare Regression*, die hier jedoch aus Platzgründen nur knapp dargestellt werden soll. In der sozialwissenschaftlichen Forschung gehört sie zu den am häufigsten verwendeten Analysemethoden (z.B. Bortz/Schuster 2010) und wird üblicherweise verwendet, um den Zusammenhang zwischen einer abhängigen Variablen (Kriteriumsvariable) mit mehreren unabhängigen Variablen (Prädiktorvariablen) sowie deren jeweiligen Anteil (Regressionsgewicht b bzw. standardisiertes Regressionsgewicht Beta) am Gesamtzusammenhang zu untersuchen. Statistisch gesehen handelt es sich auch bei der bivariaten Korrelation um eine bivariate lineare Regression.

Möchte man eine multiple lineare Regression in SPSS berechnen, erfolgt dies über die Menüpunkte *Analysieren – Regression – linear*. Im Ergebnisfenster werden dann unter anderem die Regressionskonstante a sowie die nicht-standardisierten und standardisierten Regressionsgewichte (b/Beta) ausgegeben. Des Weiteren erfolgen Angaben zur jeweiligen Irrtumswahrscheinlichkeit („Signifikanz"). Auch hier gilt: Ist dieser Signifikanzwert kleiner als 0,05, so kann davon ausgegangen werden, dass die jeweilige Prädiktorvariable unter Berücksichtigung der übrigen ins Modell aufgenommenen Variablen signifikant zur Vorhersage bzw. zur Erklärung der Kriteriumsvariablen beiträgt. Das Regressionsgewicht wiederum gibt Auskunft über die relative Wichtigkeit der gewählten Prädiktorvariablen im Rahmen des gewählten Modells. Das Regressionsgewicht (standardisiert/unstandardisiert) kann sowohl positiv als auch negativ sein. Auch hier gilt: Je näher der Wert an Null liegt, desto geringer ist der Zusammenhang.

Bezogen auf unser Beispiel könnte etwa von Interesse sein, die muttersprachliche Kompetenz (abhängige Variable oder Kriterium) mithilfe der unabhängigen Variablen (auch: Prädiktoren) „Alter in Jahren" und „Aufenthaltsdauer im Herkunftsland (in Jahren) vor der Migration" zu erklären. Für die Variable „Alter" ergibt sich kein signifikantes positives Regressionsgewicht ($p<0,05$), für die Variable „Aufenthaltsdauer im Herkunftsland" hingegen schon. In diesem Fall trägt also nur die Variable „Aufenthaltsdauer im Herkunftsland" signifikant zur Erklärung der Unterschiedlichkeit von Personen (Varianz) bezüglich der muttersprachlichen Kompetenz bei, und es gilt: je länger der Aufenthalt im Herkunftsland vor der Migration, desto besser die muttersprachliche Kompetenz (unter Kontrolle des Alters).

Wichtig ist außerdem zu beachten, dass sich die Regressionsgewichte je nach im Modell aufgenommenen Variablen ändern können. Dies ist darauf zurückzuführen, dass bei der Berechnung der linearen Regression die Zusammenhänge zwischen zwei Prädiktorvariablen „herauspartialisiert", d.h. herausgerechnet werden (dazu sowie zur Darstellung der Regressionsgleichung sowie der weitergehenden Interpretation siehe auch z.B. Rasch et al. 2006: 146ff; Bortz/Schuster 2010). Neben den Regressionsgewichten wird im Ergebnisfenster außerdem Auskunft darüber gegeben, wie gut das Gesamt-Modell „passt", d.h. wie gut die gewählten Prädiktorvariablen dazu beitragen, das Kriterium zu „erklären" oder „vorherzusagen".

Ziel ist, mithilfe der gewählten unabhängigen Variablen möglichst viel der Varianz der Kriteriumsvariablen (d.h. der Unterschiedlichkeit der Personen hinsichtlich dieses Merkmals) zu erklären. Dazu wird der sogenannte Determinationskoeffizient R^2 herangezogen. Dieser gibt Auskunft darüber, welcher Anteil der Kriteriumsvarianz in einem Regressionsmodell insgesamt erklärt werden kann. R^2 liegt zwischen 0 und 1, wobei 1 bedeuten würde, dass die Varianz der Kriteriumsvariablen perfekt anhand der Prädiktorvariablen erklärt werden kann. So könnte beispielsweise, bezogen auf unser Beispiel, ein R^2 von 0,4 so interpretiert werden, dass mithilfe der gewählten Prädiktorvariablen 40% der Kriteriumsvarianz erklärt werden kann, d.h. diese Variablen sind gemeinsam für 40% der Unterschiedlichkeit der Personen bezüglich des interessierenden Merkmals „verantwortlich".[3] Bezüglich der Interpretation aller weiteren im Ergebnisfenster dargestellten Werte sowie bezüglich der Berechnung und Interpretation der multiplen linearen Regression wird auf die oben bereits genannten Standardwerke verwiesen.

3.2.3 Überprüfung von Unterschiedshypothesen

Häufig beinhalten Forschungsfragen auch die Betrachtung von Gruppenunterschieden. Meist handelt es sich dabei um die Frage, ob sich zwei oder mehr Gruppen (im Mittel) statistisch signifikant hinsichtlich eines bestimmten Merkmals voneinander unterscheiden. Bezogen auf das oben verwendete Beispiel könnte etwa von Interesse sein, ob sich Jungen und Mädchen signifikant hinsichtlich ihrer wahrgenommenen muttersprachlichen Kompetenz unterscheiden.

---

3 Weiterführende Analysen im Hinblick auf eine Überprüfung von Kausalitäten und/oder Mediationseffekten sowie die Modellierung latenter Konstrukte können unter Anwendung fortgeschrittener statistischer Verfahren wie beispielsweise Strukturgleichungsmodellen mithilfe der Software „MPlus" (Muthén/Muthén 1998-2012) durchgeführt werden.

Ein sehr häufig für die Untersuchung von Gruppenunterschieden herangezogenes inferenzstatistisches Verfahren ist der *T-Test.* Dieser setzt Intervallskalenniveau und eine Normalverteilung der Daten voraus, ist jedoch gegenüber Verletzungen dieser Voraussetzungen gegenüber relativ robust (Bortz/Schuster 2010) und auch bereits für kleinere Stichproben geeignet. Der T-Test ist für den Vergleich zweier Gruppenmittelwerte geeignet; je nach Messdesign verwendet man den T-Test für *unabhängige* oder für *abhängige Gruppen* (bzw. Stichproben). Ersterer ist zur Betrachtung von sich gegenseitig ausschließenden Gruppen geeignet (wie z.B. Jungen und Mädchen), letzterer beispielsweise für die Betrachtung ein- und derselben Gruppe zu zwei Messzeitpunkten, beispielsweise vor und nach einer Intervention wie etwa einem Sprachtraining.

In SPSS wird der T-Test über die Menüpunkte *Analysieren – Mittelwerte vergleichen –T-Test für unabhängige/abhängige Stichproben* durchgeführt. Im folgenden Dialogfenster kann dann zunächst die Gruppierungsvariable (beispielsweise „Geschlecht", unterteilt in die beiden Gruppen „Männer" und „Frauen") ausgewählt werden. Dabei ist zu beachten, dass unter dem Button *Gruppen definieren* die verwendete Codierung angegeben werden muss (beispielsweise 1 und 2 für den Fall, dass die beiden Ausprägungen der Variablen „Geschlecht" anhand dieser Zahlenwerte codiert wurden). Im Feld *Testvariablen* werden diejenigen Variablen angegeben, die bezüglich eines möglichen Gruppenunterschieds betrachtet werden sollen.

Im Ergebnisfenster werden zunächst einige Gruppenstatistiken wie die Gruppenmittelwerte angegeben. In der zweiten Ergebnistabelle wird außerdem zuerst ein Test auf *Varianzgleichheit* (Levene-Test) der beiden Gruppen ausgegeben. Sind die Varianzen nicht gleich, ist der zum Levene-Test zugehörige Signifikanz-Wert kleiner als 0,05. Je nachdem ob die Varianzen gleich sind oder nicht, muss das Ergebnis des T-Tests in einer unterschiedlichen Zeile („Varianzen sind gleich", bzw. „Varianzen sind nicht gleich") abgelesen werden. Auch für den T-Test gilt: Ist der zugehörige Signifikanz-Wert kleiner 0,05, so unterscheiden sich die Gruppen statistisch signifikant. Ferner werden noch die mittlere Differenz der Gruppen, der dazugehörige Standardfehler sowie das Konfidenzintervall der Differenz angegeben.

Soll hingegen überprüft werden, ob sich *mehrere Gruppen* signifikant voneinander unterscheiden, und wenn ja, welche genau (bezogen auf unser Beispiel könnte etwa ein Vergleich unterschiedlicher Sprachgruppen bezüglich ihrer fremdsprachlichen Kompetenzen interessieren), dann kann diese Fragestellung anhand einer *Varianzanalyse* (sog. ANOVA: *An*alysis *o*f *Va*riance) bearbeitet werden. Aufgrund der Komplexität der Analysen kann die Varianzanalyse hier lediglich rudimentär dargestellt werden; bezüglich mathematischer Grundlagen

und einer weitergehenden Ergebnisinterpretation wird auf die entsprechende Fachliteratur (z.B. Bortz/Schuster 2010; Janssen/Laatz 2013) verwiesen. mithilfe der einfaktoriellen Varianzanalyse kann beispielsweise untersucht werden, ob sich Gruppen (in SPSS bezeichnet als Faktor) hinsichtlich eines Merkmals (abhängige Variable) unterscheiden. Dazu wird in einem ersten Schritt zunächst untersucht, ob über alle Gruppen hinweg ein Unterschied zu finden ist.

In SPSS ist die Varianzanalyse unter den Menüpunkten *Mittelwerte vergleichen – einfaktorielle ANOVA* zu finden. Im folgenden Dialogfenster muss nun angegeben werden, welche Variable die zu betrachtenden Gruppen codiert und beinhaltet (Faktor), und welche Variablen auf Gruppenunterschiede hin untersucht werden sollen (abhängige Variable). Im Ausgabefenster ist nun die Tabelle „einfaktorielle ANOVA“ einzusehen. Diese beinhaltet neben Angaben zu den Quadratsummen innerhalb und zwischen den Gruppen und der Prüfgröße (dazu z.B. Bortz/Schuster, 2010) auch einen dazugehörigen Signifikanzwert, der zu interpretieren ist. Ist dieser Wert signifikant, dann ist dies ein Hinweis darauf, dass über alle Gruppen hinweg ein statistisch signifikanter Unterschied zu finden ist.

Um welche Gruppen genau es sich dabei handelt, wird erst im Anschluss mithilfe sogenannter *Post-hoc-Tests* analysiert. Je nachdem ob Varianzgleichheit angenommen wird oder nicht (ein *Test auf Varianzhomogenität* ist zu finden unter dem Button „Optionen“) und in Abhängigkeit davon, ob die Stichprobenumfänge gleich sind, existieren unter dem Button *Post hoc* verschiedene Auswahlmöglichkeiten. In der Ergebnis-Tabelle *Mehrfachvergleiche* ist dann einzusehen, zwischen welchen der Gruppen genau sich statistisch signifikante Unterschiede finden. Neben dem dafür relevanten Signifikanzwert finden sich in der Tabelle außerdem Angaben zur mittleren Differenz der Gruppen, Standardfehler und Konfidenzintervall.

3.2.4 Weitere Analysen

Neben den im Rahmen des vorliegenden Kapitels dargestellten Verfahren kann mithilfe von SPSS noch eine Vielzahl weiterer Analysen durchgeführt werden. Diese beinhalten beispielsweise unter dem Punkt „Skalierung“ die Möglichkeit, die Reliabilität von Skalen zu berechnen. Auch können Analysen im Rahmen des Allgemeinen Linearen Modells oder Verfahren zur Dimensionsreduzierung wie die explorative Faktorenanalyse durchgeführt werden. Eine ausführliche Beschreibung aller weiteren, in SPSS möglichen Analysen inklusive der Interpretation von Ergebnisoutputs findet sich etwa bei Janssen/Laatz (2013) und Brosius (2013).

Insgesamt wird deutlich, dass eine gemeinsame Analyse quantitativer und qualitativer Daten mithilfe computerbasierter Auswertungsprogramme wie SPSS möglich ist. Dies erfordert jedoch eine vorausgehende Quantifizierung der qualitativen Daten. Das dazu notwendige Vorgehen wurde im Rahmen des Kapitels dargestellt. Ist dies erfolgt, können für alle Daten, unabhängig davon, ob diese mit qualitativen oder quantitativen Befragungsmethoden gewonnen wurden, prinzipiell die gleichen Analyseverfahren angewandt werden. Dies ermöglicht eine gemeinsame Interpretation der Daten und führt zu einer simultanen Verwendbarkeit qualitativer und quantitativer Befragungsmethoden.

Literatur

Bortz, Jürgen / Schuster, Christoph (2010). *Statistik für Human- und Sozialwissenschaftler*. 7. Aufl. Berlin Heidelberg: Springer.

Borg, Ingwer / Stauffenbiel, Thomas (1997). *Theorien und Methoden der Skalierung. Eine Einführung*. Methoden der Psychologie, Band 11. 3. Aufl. Bern: Huber.

Breuer, Franz (2010). Wissenschaftstheoretische Grundlagen qualitativer Methodik in der Psychologie. In: Mey, Günter / Mruck, Katja (Hrsg.). *Handbuch Qualitative Forschung in der Psychologie.* Wiesbaden: VS Verlag für Sozialwissenschaften, 35-45.

Broda, Stephan (2006). *Marktforschungs-Praxis. Konzepte, Methoden, Erfahrungen*. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler/GWV Fachverlage.

Brosius, Felix (1998). *SPSS 8. Professionelle Statistik unter Windows*. Mitp-Verlag.

Brosius, Felix (2013). *SPSS 21 (mit Professional)*. Mitp-Verlag.

Bühner, Markus (2006). *Einführung in die Test- und Fragebogenkonstruktion*. 2. Aufl. München: Pearson Studium.

Flick, Uwe (2011). *Triangulation. Eine Einführung.* 3. Aufl. Reihe Qualitative Sozialforschung, Band 12. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gläser-Zikuda, Michaela / Seidel, Tina / Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha (2012). Mixed Methods in der empirischen Bildungsforschung – eine Einführung in die Thematik. In: Gläser-Zikuda, Michaela / Seidel, Tina / Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha (Hrsg.). *Mixed Methods in der empirischen Bildungsforschung*. Münster: Waxmann, 7-14.

Janssen, Jürgen / Laatz, Wilfried (2013). *Statistische Datenanalyse mit SPSS. Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests*. 8. Aufl. Berlin, Heidelberg: Springer.

Jonkisz, Ewa / Moosbrugger, Helfried / Brand, Holger (2011). Planung und Entwicklung von Tests und Fragebogen. In: Moosbrugger, Helfried / Kelava, Augustin (Hrsg.). *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer, 27-74.

Kelle, Udo (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kempf, Wilhelm (2010). *Diskussionsbeiträge der Projektgruppe Friedensforschung Konstanz*, Nr. 65. Berlin: Regener. Online: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-121212 [10.3.2014].

Konrad, Klaus (2011). *Mündliche und schriftliche Befragung – Ein Lehrbuch*. Forschung, Statistik und Methoden, Band 4. 7. Aufl. Landau: Verlag Empirische Pädagogik.

Mayring, Philipp (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* 10. Aufl.. Weinheim: Beltz.

Mayring, Philipp (2011). Qualitative Inhaltsanalyse. In: Mey, Günter / Mruck, Katja (Hrsg.). *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften, 601-613.

Messick, Samuel (1989). Validity. In: Linn, Robert L. (Hrsg.). *Educational Measurement*. 3. Aufl. New York: Macmillan.

Mey, Günter / Mruck, Katja (2011). Interviews. In: Mey, Günter / Mruck, Katja (Hrsg.). *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften, 423-436.

Muthén, Linda K. / Muthén, Bengt O. (1998-2012). *Mplus User's Guide*. 7. Aufl. Los Angeles, C.A.: Muthén & Muthén.

The R Foundation for Statistical Computing (2014). „*R*". Online: http://www.r-project.org [10.3.2014].

Rasch, Björn / Friese, Malte / Hofmann, Wilhelm / Naumann, Ewald (2006). *Quantitative Methoden. Einführung in die Statistik*. Heidelberg: Springer.

Rost, Jürgen (1996*). Testtheorie, Testkonstruktion*. Bern: Verlag Hans Huber.

Schmidt-Atzert, Lothar / Amelang, Manfred (2012). *Psychologische Diagnostik*. Berlin, Heidelberg: Springer.

Settinieri, Julia (2014). Forschst Du noch, oder triangulierst Du schon? In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 17-35.

SPSS inc./IBM (2013). *IBM SPSS Statistics 21*. Online: http://www-01.ibm.com/software/de/analytics/spss/ [10.3.2014].

Strauss, Anselm / Corbin, Juliet (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz.

Competences and Motivation in the Bilingual History Classroom (COMBIH) – Eine Längsschnittstudie zu deutsch-englischem Geschichtsunterricht an Gymnasien

Sara Dallinger[1]*/ Kathrin Jonkmann*

Die Studie COMBIH untersucht die Entwicklung der Englisch- und Geschichtsleistungen sowie der Motivation und Emotionen der Schülerinnen und Schüler im bilingualen Geschichtsunterricht. Dabei werden Eingangsunterschiede zwischen bilingual und regulär einsprachig unterrichteten Lernern zu Beginn des bilingualen Geschichtsunterrichts berücksichtigt. Der Schwerpunkt des Beitrags liegt auf der Darstellung und Erörterung des triangulativen Gehalts der quantitativen und längsschnittlichen Studie, bei der zu zwei Messzeitpunkten in Klassenstufe 8 Kompetenztests und Fragebögen für Schülerinnen und Schüler sowie Lehrkräfte an Gymnasien in Baden-Württemberg eingesetzt wurden.

1. Einleitung

In den letzten Jahren kann in einigen Bundesländern ein Trend zu einer verstärkten und flächendeckenden Einführung bilingualen Sachfachunterrichts beobachtet werden. Beispielsweise wurde in Baden-Württemberg an 65 Gymnasien ein bilingualer Zug eingerichtet und seit 2010 wird deutsch-englischer Sachfachunterricht auch an bis zu 20% aller Realschulen etabliert. In Nordrhein-Westfalen wurde das Programm „Bili für alle" entwickelt (vgl. Breidbach/Viebrock 2012: 6). Diese politischen Entwicklungen stützen sich teilweise auf wissenschaftliche Studien, die zu überwiegend positiven Ergebnissen bezüglich des bilingualen Sachfachunterrichts kamen: Bilingual unterrichtete Schülerinnen und Schüler scheinen sowohl hinsichtlich ihrer Englisch- als auch ihrer Sachfachkompetenzen von dieser Unterrichtsform zu profitieren (Zydatiß 2007, Lamsfuß-Schenk 2008, Nold et al. 2008,Witzigmann 2011). Weiterhin zeigen sich sowohl Schulleitungen (Hollm et al. 2013), Lehrkräfte (Kollenrott 2008) als auch Eltern

1 Sara Dallinger ist Mitglied des Kooperativen Promotionskollegs „Effektive Lehr-Lernarrangements" der Pädagogischen Hochschule Ludwigsburg und der Universität Tübingen, das vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert wird.

(Dallinger 2013) weitgehend überzeugt vom bilingualen Sachfachunterricht. Im Gegensatz zu dieser „Erfolgsgeschichte" des bilingualen Lehrens und Lernens stehen Arbeiten, die diese in Frage stellen. Bruton (2011) verweist auf die in vielen Studien fehlende Analyse der Ausgangslage von bilingual und einsprachig unterrichteten Schülerinnen und Schülern, wonach bilingual unterrichtete Lerner meist schon zu Beginn des bilingualen Unterrichts die leistungsstärkeren und motivierteren Schülerinnen und Schüler einer Schule sind. Ihre weiterhin positive Entwicklung könnte daher auf diese Eingangsunterschiede statt auf den bilingualen Unterricht zurückgeführt werden (Zydatiß 2012: 23). Dies führt vor allem bei querschnittlichen Studien zu einer systematischen Überschätzung des Effekts bilingualen Unterrichts (Rumlich 2014). So argumentiert auch Bonnet: „There are strong indications from evaluation research […] that social selection plays an important role in German CLIL. Therefore, high outcomes in CLIL groups might well be ascribed to factors other than classroom practice" (2012: 67). Es werden also Längsschnittstudien benötigt, die die unterschiedlichen Lernvoraussetzungen und Ausgangslagen von bilingual und einsprachig unterrichteten Schülerinnen und Schülern berücksichtigen und diese bei der Vorhersage von Veränderungen entsprechend kontrollieren, um somit den Effekt des bilingualen Unterrichts auf Schülerkompetenzen und Motivation bestmöglich isolieren zu können (Costa/D'Angelo 2011: 3). Die diesem Beitrag zugrunde liegende Studie COMBIH möchte versuchen, diese Forschungslücken zu schließen.

2. Längsschnittliche Forschung zum bilingualen Sachfachunterricht

Die Unterschiede in den Lernausgangslagen bilingual und einsprachig unterrichteter Schülerinnen und Schüler gehen großteils auf die Auswahlprozesse zurück, die an Schulen mit fest eingerichtetem bilingualen Profil stattfinden. Diese Schulen besitzen neben anderen Spezialisierungen (wie z. B. musisches oder naturwissenschaftliches Profil) einen bilingualen Zug, den Schülerinnen und Schüler bzw. Eltern am Ende der Grundschule wählen können. Die Auswahl seitens der Schulverwaltung basiert einerseits auf den Grundschulnoten und Empfehlungsschreiben der Grundschullehrkräfte. Anderseits stellen Gespräche zwischen Schulverwaltung und Eltern einen weiteren zentralen Baustein des Auswahlverfahrens dar (Zydatiß 2007: 70). Da in diesen Gesprächen häufig vermittelt wird, dass bilingualer Unterricht mit höheren kognitiven Ansprüchen und Lernaufwand verbunden ist (Hollm et al. 2010: 5, Rumlich 2013a: 184), darf vermutet werden, dass sich vor allem besonders motivierte und interessierte Kinder mit hoher Lernbereitschaft für den bilingualen Zug bewerben.

In diesem Sinne kann von einer positiv selektierten Schülerschaft gesprochen werden, wenn diese bessere Lernvoraussetzungen aufweist als eine Vergleichsgruppe, wie etwa höhere kognitive Grundfähigkeiten (Helmke/Schrader 2001: 81), eine höhere Leistungsbereitschaft und Motivation (Wigfield/Eccles 2000) und eine größere Bildungsnähe des Elternhauses (Köller et al. 2012: 339). Über diese allgemeinen Lernvoraussetzungen hinaus ist es für Studien, die erst im Laufe des bilingualen Zuges einsetzen, wichtig, die Unterschiede in den Englischleistungen und das Vorwissen im Sachfach zu erfassen und zu kontrollieren, um den inkrementellen Effekt des bilingualen Sachfachlernens im Laufe eines Schuljahres ermitteln zu können. Im Folgenden wird nun eine Auswahl an Studien präsentiert, die aufgrund ihres Längsschnittdesigns diese Ausgangslagen der bilingual und einsprachig unterrichteten Schülerschaften berücksichtigen.

Mit C-Tests verglich Rumlich (2013a) in seiner Studie DENOCS die allgemeine Englischkompetenz einer Teilstichprobe von 340 bilingual und 430 nicht bilingual unterrichteten Gymnasialschülern. Die nicht bilingual unterrichtete Stichprobe setzte sich aus zwei Untergruppen zusammen: Schülerinnen und Schüler aus den Parallelklassen der bilingual unterrichteten Klassen sowie Schülerinnen und Schüler von Schulen ohne bilinguales Unterrichtsangebot. Die Ergebnisse der ersten Testung am Ende der Klasse 6, vor Beginn des bilingualen Sachfachunterrichts, verweisen auf deutlich höhere Englischleistungen der zukünftig bilingual unterrichteten Lerner. Bezüglich der beiden Vergleichsgruppen schnitten die Schülerinnen und Schüler aus Schulen ohne bilinguales Angebot besser ab als die Parallelklassenschüler. Dies deutet darauf hin, dass innerhalb einer Schule eher die leistungsstärkeren Schülerinnen und Schüler den bilingualen Zug wählen. Weiterhin wurde ein Fragebogen eingesetzt, der unter anderem das Schülerinteresse an der Fremdsprache und am Fach Englisch erfasste. Auch hier wurde festgestellt, dass bilingual unterrichtete Lerner mehr Interesse zeigten als die beiden nicht bilingual unterrichteten Schülergruppen (Rumlich 2014). Die Ergebnisse der zweiten Messung zum Ende der Klasse 7 verweisen darauf, dass der Leistungsvorsprung der bilingual unterrichteten Schülerinnen und Schüler vom ersten Messzeitpunkt bezüglich der allgemeinen Englischkompetenz zwar bestehen blieb, sich dieser aber im Verlauf eines Schuljahres nicht weiter vergrößerte (Rumlich 2013b). Bilingual und einsprachig unterrichtete Lerner wiesen damit weitgehend parallel verlaufende Leistungszuwächse auf; bei den bilingual unterrichteten Schülerinnen und Schülern erfolgte dies lediglich auf einem höheren Niveau.

Fehling (2008) untersuchte in ihrer Studie die *Language Awareness*. Zum ersten Messzeitpunkt, zu Beginn Klasse 7, erhob sie Daten von 74 bilingual und 65 nicht bilingual unterrichteten Gymnasialschülern. Mithilfe des „*Language*

Awareness-Tests" und einem Test zur Erfassung der verbalen kognitiven Grundfähigkeiten konnte Fehling zeigen, dass bilingual unterrichtete Schülerinnen und Schüler in drei von vier gemessenen Dimensionen der *Language Awareness* (affektiv, kognitiv, politisch) bessere Resultate erzielten. Auch beim kognitiven Grundfähigkeitstest schnitten die bilingual unterrichteten Lerner besser ab. Im Verlauf der Klassenstufen 7 und 8 zeigten bilingual unterrichtete Schülerinnen und Schüler anschließend größere Zuwächse bezüglich der affektiven und kognitiven Dimension als einsprachig unterrichtete Lerner. Das Interesse an Englisch blieb bei den bilingual unterrichteten Schülerinnen und Schülern auf demselben Niveau wie zu Beginn des bilingualen Unterrichts. Bei den einsprachig unterrichteten Schülerinnen und Schülern nahm das Interesse an Englisch jedoch ab. Ein Absinken des Fachinteresses im Laufe der Sekundarstufe ist ein vielfach repliziertes entwicklungspsychologisches Phänomen (Wild/Hofer 2000, Krapp 2002). Das Entgegenwirken zu diesem Trend kann durchaus als Erfolg gewertet werden.

Zur Entwicklung von Kompetenzen im bilingual unterrichteten Sachfach liegen zurzeit kaum Studien vor (Coyle et al. 2010: 136ff., Bonnet 2012: 67). Badertscher/Bieri (2009) filmten in sechs Klassen einsprachig sowie bilingual unterrichtete Unterrichtseinheiten in den Fächern Geographie und Geschichte und führten vor und nach diesen Unterrichtsstunden Interviews mit den Schülerinnen und Schülern durch, um die Wissensstände zu erfassen. Dieses Vorgehen wurde nach 1,5 Jahren in den Klassen wiederholt. Das Forscherteam kam zu dem Schluss, dass keine Unterschiede zwischen den Schülergruppen bezüglich der Qualität des von den Lernern reproduzierten Sachfachwissens bestanden. Zu diesem Ergebnis kamen auch Alonso et al. (2008), die die Fremdsprachen- und Sachfachkompetenzen von 229 bilingual und einsprachig unterrichteten Sekundarstufenschülern im Baskenland untersuchten. Neben einem größeren Leistungszuwachs der bilingual unterrichteten Lerner in Bezug auf den Cambridge ESOL Test resümierte das Forscherteam, dass diese Schülergruppe ähnliche Leistungen wie die einsprachig unterrichteten Lerner in den Sachfächern zeigte. Diese Aussage stützte sich jedoch lediglich auf die subjektive Einschätzung von Lehrkräften sowie auf die Noten der Schülerinnen und Schüler, die Referenzgruppeneffekten unterliegen (Baumert et al. 2003). Zusammenfassend gilt anzumerken, dass in diesen Studien keine standardisierten und objektiven Leistungstests zur Erfassung des Sachfachwissens eingesetzt wurden.

Auch in Bezug auf die Effekte bilingualen Unterrichts auf die Veränderung der Motivation im Sachfach und in der Fremdsprache ist die empirische Datenbasis recht klein (Lasagabaster/Sierra 2009: 4). Abendroth-Timmer (2004) untersuchte bilingual und mehrsprachig unterrichtete Module der Sekundarstufe

in verschiedenen Sachfächern für die Arbeitssprachen Englisch, Französisch und Spanisch. Mithilfe von Einzelfallanalysen, die auf Schülerinterviews basierten, und den Daten von Schülerfragebögen kam sie zu dem Schluss, dass bilingualer Unterricht für die Motivation der Schülerinnen und Schüler nicht abträglich sei, dies jedoch von verschiedenen Faktoren abhänge. So kann bilingualer Unterricht bei angemessenem Sprach- und inhaltlichem Niveau sowie unter Berücksichtigung sachfachdidaktischer Prinzipien seitens der Lehrkraft (z. B. Authentizität, Handlungsbezug) förderlich auf die Motivation wirken. Zahlreiche querschnittliche Studien bestätigen dieses Ergebnis (z. B. Zydatiß 2007, Lamsfuß-Schenk 2008). Diese Befunde sollten jedoch mittels standardisierter Fragebögen zur differenzierten Erfassung unterschiedlicher motivationaler Konstrukte wie z. B. Fähigkeitsüberzeugungen, Fachinteresse, Nutzenbewertung oder Leistungsemotionen im Rahmen längsschnittlicher Studien auf eine breitere empirische Basis gestellt werden.

Abschließend werden zwei Studien querschnittlichen Designs vorgestellt, die zwar nicht die Ausgangslagen in Englisch, dafür aber zentrale Indikatoren der Lernvoraussetzungen bilingual und einsprachig unterrichteter Schülerinnen und Schüler berücksichtigten. Nold et al. (2008) untersuchten mithilfe einer Teilstichprobe von DESI die Englischkompetenzen von rund 2000 Gymnasialschülern. Hierbei wurden die Daten des zweiten Messzeitpunkts (Ende der 9. Klassenstufe) herangezogen. Um dem Problem der positiveren Ausgangslage bilingual unterrichteter Lerner zu begegnen, wurden die Schülerinnen und Schüler der Vergleichsgruppe gezielt ausgewählt. Es wurden 987 Lerner bestimmt, die sich auf Basis der Daten des ersten Messzeitpunkts (Beginn der 9. Klassenstufe) nicht von den bilingual unterrichteten Schülerinnen und Schülern unterschieden hinsichtlich ihres Geschlechts, der Erstsprache, Schulabschluss und Beruf der Eltern, der Note in Deutsch und den kognitiven Grundfähigkeiten. Die Ergebnisse zeigten, dass bilingual unterrichtete Schülerinnen und Schüler in vier Kompetenzbereichen (Hör- und Leseverstehen, Grammatik, Textrekonstruktion) das jeweils höhere Kompetenzniveau in Bezug auf den Gemeinsamen Europäischen Referenzrahmen für Sprachen erreichten als die einsprachig unterrichteten Lerner. Die größten Leistungsvorteile traten im Bereich des Hörverstehens auf (DESI-Konsortium 2006: 59). Köller et al. (2012) konnten diese Befunde replizieren. Sie analysierten die Daten einer Teilstichprobe von 345 bilingual und 9521 einsprachig unterrichteten Schülerinnen und Schüler der 9. Klassenstufe bezüglich ihrer Englischkompetenzen (Hör- und Leseverstehen). Folgende Hintergrundmerkmale wurden statistisch kontrolliert, um die Unterschiedlichkeit der beiden Schülergruppen zu berücksichtigen: Geschlecht, Erstsprache, Migrationshintergrund, sozioökonomischer Hintergrund, Beruf der Eltern, Deutschkompetenzen der Lerner. Trotz Kontrolle dieser Einflüsse auf die Englischleistungen zeigte

sich, dass die bilingual unterrichteten Schülerinnen und Schüler deutlich besser in den Tests abschnitten als die regulär einsprachig unterrichteten Lerner.

Zusammenfassend scheinen bilingual unterrichtete Schülerinnen und Schüler eine Schülergruppe zu sein, die über besonders gute Lernvoraussetzungen verfügt. Diese Unterschiede müssen Studien, die die Effekte bilingualen Unterrichts bestimmen möchten, kontrollieren. Empirische Befunde weisen beispielsweise darauf hin, dass der Leistungsvorsprung in Englisch auch dann noch besteht, wenn zentrale Hintergrundmerkmale (z. B. kognitive Grundfähigkeiten, Bildungsnähe des Elternhauses) kontrolliert werden. Ob diese Unterschiede jedoch auf den erhöhten Stundenumfang in Englisch in den Klassenstufen 5 und 6, der von Schülerinnen und Schülern eines bilingualen Zuges in der Regel geleistet werden muss, zurückgehen, lassen die Daten der zitierten Studien offen, da die erste Messung zwar immer zu Beginn des bilingualen Sachfachunterrichts, jedoch nicht schon zu Beginn des verstärkten Englischunterrichts durchgeführt wurde. Weiterhin deuten Rumlichs Ergebnisse an (2013b), dass durch den bilingualen Sachfachunterricht keine inkrementellen Zuwächse erzielt werden, während Fehling (2008) für Englisch größere Leistungszuwächse bilingual unterrichteter Lerner feststellt. Auch in Bezug auf die Entwicklung von Sachfachkompetenzen sowie von Motivation und Emotionen liegen bisher nur wenige und zum Teil widersprüchliche Ergebnisse vor. Es bedarf an weiterer, belastbarer empirischer Forschung zur Validierung und Ausdifferenzierung dieser Befunde, die nicht nur zentrale Aspekte wie Kompetenzen und Motivation von Schülerinnen und Schülern im bilingualen Unterricht in den Blick nimmt, sondern auch wichtige Kontextvariablen und Hintergrundmerkmale berücksichtigt (Pérez-Cañado 2012: 330ff.).

3. Forschungsfragen

Für die COMBIH-Studie ergibt sich aus den dargestellten Befunden folgender Forschungsschwerpunkt: Wie wirkt sich bilingualer Geschichtsunterricht auf die Leistungsentwicklung, Motivation und Emotionen der Schülerinnen und Schüler in den Fächern Englisch und Geschichte aus? Zur Beantwortung dieser Frage wird ein längsschnittliches Untersuchungsdesign mit einem ersten Messzeitpunkt zu Beginn des bilingualen Geschichtsunterrichts in Klassenstufe 8 durchgeführt, um Unterschiede in den Ausgangslagen zwischen den Schülerschaften und gegebenenfalls zwischen den Lehrerschaften kontrollieren und somit potenzielle differentielle Veränderungen im Laufe der Klasse 8 bestmöglich auf den bilingualen Unterricht zurückführen zu können. Selbstverständlich ist auch bei diesem Untersuchungsdesign die Aussagekraft über Ursache-Wirkungs-Zusammenhänge eingeschränkt und davon abhängig, ob die

zentralen konfundierenden Variablen gemessen und kontrolliert werden. Echte Kausalaussagen sind nur mit einem experimentellen Design und randomisierter Bedingungszuweisung möglich, bei dem Störvariablen und Alternativerklärungen ausgeschlossen werden können. Da dies in Bildungssettings meist nicht realisierbar ist, muss mit Längsschnittstudien und der Verwendung geeigneter statistischer Verfahren eine bestmögliche Approximation geleistet werden. Die Forschungsfragen von COMBIH lauten wie folgt:

Lernvoraussetzungen und Ausgangslagen von bilingual und einsprachig unterrichteten Schülerinnen und Schülern zu Beginn der 8. Klassenstufe

- Inwiefern unterscheiden sich bilingual und einsprachig unterrichtete Schülerinnen und Schüler zu Beginn des bilingualen Geschichtsunterrichts hinsichtlich ihres familiären Hintergrunds, ihrer kognitiven Grundfähigkeiten, ihrer Schulnoten sowie ihrer Leistungen, Motivation und Leistungsemotionen in den Fächern Englisch und Geschichte?
- Unterscheiden sich die Englisch- und Geschichtsleistungen von bilingual und einsprachig unterrichteten Schülerinnen und Schülern auch dann noch, wenn zentrale Hintergrundmerkmale kontrolliert werden?
- Gibt es Unterschiede zwischen den beiden nicht bilingual unterrichteten Vergleichsgruppen hinsichtlich der zentralen Konstrukte?

In Anlehnung an die berichteten Ergebnisse zu den Ausgangslagen der beiden Schülergruppen wird erwartet, dass die bilingual unterrichteten Schülerinnen und Schüler in den Leistungstests und Schulnoten besser abschneiden als die einsprachig unterrichteten Lerner. Außerdem werden für beide Fächer positivere Motivationswerte und Emotionen bei den bilingual Unterrichteten erwartet. Darüber hinaus wird angenommen, dass der Leistungsvorsprung der bilingualen Lerner in Englisch auch dann noch besteht, wenn Hintergrundmerkmale kontrolliert werden. Die Leistungsvorsprünge in Geschichte hingegen sollten weitgehend durch die Unterschiede in den kognitiven Grundfähigkeiten und im familiären Hintergrund erklärt werden. Weiterhin wird angenommen, dass die Schülerinnen und Schüler der nicht bilingual unterrichteten Parallelklassen niedrigere Leistungen und Motivation aufweisen als die Schülerinnen und Schüler von Schulen ohne bilinguales Unterrichtsangebot.

Veränderungen von Kompetenzen, Motivation und Emotionen in der 8. Klassenstufe in Abhängigkeit vom bilingualen Unterricht

- Können bilingual unterrichtete Schülerinnen und Schüler einen möglichen Leistungsvorsprung gegenüber einsprachig unterrichteten Schülerinnen und

Schülern in den Fächern Englisch und Geschichte im Lauf eines Jahres bilingualen Geschichtsunterrichts weiter ausbauen?
- Verändern sich die Motivation und die Leistungsemotionen der bilingual unterrichteten Schülerinnen und Schüler in den Fächern Englisch und Geschichte positiver als bei einsprachig unterrichteten Schülerinnen und Schülern?
- Bestehen diese differentiellen Verläufe auch dann noch, wenn die Ausgangslagen, die Hintergrundmerkmale und Kontextvariablen kontrolliert werden?

Bezüglich der Englischkompetenzen wird erwartet, dass die bilingual unterrichteten Schülerinnen und Schüler einen größeren Leistungszuwachs erfahren als die einsprachig unterrichteten Lerner, und dass diese Unterschiede auch nach Kontrolle der Ausgangslagen, der Hintergrundmerkmale und Kontextvariablen statistisch bedeutsam bleiben. Wie sich die Kompetenzen im Fach Geschichte entwickeln, ist eine offene Forschungsfrage. Bezüglich der Motivation und Emotionen bilingual unterrichteter Schülerinnen und Schüler kann erwartet werden, dass diese auf demselben (erhöhten) Niveau bleiben wie beim ersten Messzeitpunkt, während für die einsprachig unterrichteten Schülerinnen und Schüler der typische Entwicklungsverlauf eines sinkenden Interesses vorliegt.

4. Design der Studie

4.1 Stichprobe

Die Stichprobe der Studie bestand aus Schülerinnen und Schülern der 8. Klassenstufe von Gymnasien in Baden-Württemberg. Sie gliederte sich in drei Untersuchungsgruppen auf. Die erste Gruppe setzt sich aus Schülerinnen und Schülern zusammen, die ab Klasse 5 den bilingualen Zug ihrer Schule besuchten. In den Klassenstufen 5 und 6 nahmen sie am verstärkten Englischunterricht teil (2 zusätzliche Schulstunden pro Woche), und ab Klasse 7 wurden sie in mindestens einem Fach pro Schuljahr durchgängig bilingual unterrichtet. In Klasse 7 war dies Geographie, in der 8. Klasse nahmen die Schülerinnen und Schüler zum ersten Mal an bilingualem Geschichtsunterricht teil. Regulär deutschsprachiger Geschichtsunterricht wurde bereits ab Klasse 6 erteilt. Die Schülerinnen und Schüler der Vergleichsgruppe I besuchten die Parallelklassen der bilingualen Gruppen, d. h. sie hatten sich gegen den bilingualen Zug entschieden oder wurden in diesen nicht aufgenommen. Vergleichsgruppe II umfasste Schülerinnen und Schüler von Gymnasien ohne bilinguales Unterrichtsangebot. Die Lehrerstichprobe setzte sich aus den Geschichtslehrkräften der an der Studie

beteiligten Klassen zusammen. Um eine hinreichend hohe statistische Teststärke zu gewährleisten, war es das Ziel der Studie, mindestens 20 Klassen pro Gruppe zu rekrutieren. Um dies zu erreichen, wurden alle 65 Gymnasien mit bilingualem Zug in Baden-Württemberg gezielt kontaktiert und um Teilnahme gebeten. Die Schulen der Vergleichsgruppe II wurden aus Machbarkeitsgründen aus dem näheren Umkreis der Hochschulen gewählt. 28 der Gymnasien mit bilingualem Zug nahmen an der Studie teil sowie 9 weitere Gymnasien ohne bilinguales Angebot (vgl. Tab. 1). An der Vorstudie, die der Pilotierung von Teilen der Erhebungsmaterialien diente, nahmen 12 Klassen aus vier Gymnasien mit 222 Schülerinnen und Schülern der Klassenstufen 7 und 8 teil.

Tab. 1: Stichprobe der Hauptstudie (MZP = Messzeitpunkt)

Gymnasien		**Klassen**		**Schüler MZP1**	**Schüler MZP2**
mit bilingualem Zug	28	Bilingual-Klassen	30	664	661
		Vergleichsgruppe I	28	588	577
ohne bilinguales Angebot	9	Vergleichsgruppe II	17	411	394
Gesamt	**37**		**75**	**1663**	**1632**

4.2 Organisation und Ablauf der Studie

Vor Beginn der Datenerhebungen wurden alle eingesetzten Materialien und Prozeduren für eine datenschutzrechtliche Prüfung an das Ministerium für Kultus, Jugend und Sport Baden-Württemberg geleitet. Diese wurden dort begutachtet und für die Untersuchung freigegeben. Der erste Messzeitpunkt fand im September und Oktober 2012 statt. Dieser Zeitpunkt wurde gewählt, um die Schülerinnen und Schüler und Lehrkräfte zu Beginn des Schuljahres befragen zu können. Dieselbe Schüler- und Lehrerstichprobe wurde im Juni und Juli 2013, zum Ende des Schuljahres, ein zweites Mal befragt. Zum ersten Messzeitpunkt nahm die Datenerhebung insgesamt 135 Minuten in Anspruch (exklusive der Pausenzeiten), zur zweiten Messung 115 Minuten. Die Erhebungen wurden mithilfe eines standardisierten Manuals von geschulten Testleitern durchgeführt. Zusätzlich waren Lehrkräfte im Klassenzimmer anwesend. Das Test- und Befragungsmaterial der Schülerinnen und Schüler wurde vor Ort ausgeteilt und nach Beendigung der Erhebung wieder von den Testleitern eingesammelt. Die Lehrerfragebögen, die ca. 20 Minuten in Anspruch nahmen, wurden entweder an der Schule oder zuhause ausgefüllt.

4.3 Erhebungsinstrumente

Die Untersuchungsmaterialien der Schülerinnen und Schüler umfassten Leistungstests und einen Fragebogen. Die Englischleistungen wurden mittels eines eigens entwickelten C-Tests sowie mit einem Hörverstehenstest aus dem Aufgabenpool der Bildungsstandards (IQB) geprüft. Das Geschichtswissen der Schülerinnen und Schüler aus den Klassen 7 und 8 wurde mithilfe von Lückentexten und *Multiple Choice*-Aufgaben gemessen, die in Kooperation mit Geschichtsdidaktikern entwickelt wurden. Außerdem wurden Fragen zur Erfassung der historischen De- und Rekonstruktionskompetenz (Bertram 2012) in Anlehnung an das Kompetenz-Strukturmodell für historisches Denken nach Schreiber et al. (2006) integriert. Die figuralen und verbalen kognitiven Grundfähigkeiten der Schülerinnen und Schüler wurden mit Subtests des Kognitiven Fähigkeitstests erhoben (KFT, Heller/Perleth 2000). Die Fragebogenitems zur Schülermotivation in den Fächern Englisch und Geschichte orientierten sich an dem Erwartungs-Mal-Wert-Modell nach Wigfield/Eccles (2000), das sich bereits in vielen pädagogisch-psychologischen Studien bewährt hat. In diesem Modell werden unterschiedliche motivationale Konstrukte zur Erklärung leistungsbezogenen Verhaltens berücksichtigt, wie beispielsweise Interesse, subjektiver Nutzen und akademisches Selbstkonzept. Die Erfassung fachspezifischer Leistungsemotionen (z.B. Freude, Langeweile) erfolgte auf Basis der Kontroll-Wert-Theorie nach Pekrun (2006,vgl. auch Götz et al. 2010, Götz/Frenzel 2010). Im Schülerfragebogen befanden sich weiterhin Fragen zum bilingualen Geschichtsunterricht (z. B. Bewertung der Unterrichtsform, Rolle des Englischen) sowie allgemeine Fragen zur Unterrichtsqualität in den Fächern Englisch und Geschichte (COACTIV-Studie; Kunter et al. 2011). Die Operationalisierung des familiären Hintergrunds orientierte sich an den PISA-Studien. Außerdem wurden die Schulnoten der Lerner erhoben. Die Geschichtslehrkräfte der betreffenden Klassen erhielten einen Fragebogen mit eigens entwickelten Fragen zu methodisch-didaktischen Aspekten des (bilingualen) Geschichtsunterrichts, ihrer Lehrerbiographie sowie zu arbeitsbezogenen Verhaltens- und Erlebensmustern (AVEM; Schaarschmidt/Fischer 2008).

4.4 Datenanalyse

Die Studie wird derzeit unter Verwendung der Softwarepakete SPSS und Mplus insbesondere mit regressionsanalytischen Verfahren ausgewertet. Da ganze Klassen in die Untersuchung aufgenommen wurden, liegen keine vollständig unabhängigen Beobachtungen vor. Diese geschachtelte Datenstruktur wird bei den Analysen berücksichtigt. Andernfalls würden die Standardfehler unterschätzt

und es erfolgt eine zu liberale statistische Signifikanzprüfung (Raudenbush/Bryk 2002). Darüber hinaus sollen Mehrebenenregressionsmodelle spezifiziert werden (Snijders/Bosker 2011), um die Bedingungsfaktoren eines effektiven bilingualen Geschichtsunterrichts auf Klassenebene identifizieren zu können.

5. Aspekte der Triangulation

5.1 Theoretische Vorüberlegungen

Der Begriff der Triangulation wurde in den letzten Jahren verstärkt in der empirischen Forschung sowie im Rahmen definitionstheoretischer Reflexionen verwandt (Riemer 2011, Gläser-Zikuda et al. 2012). Während Einigkeit darüber herrscht, dass Triangulation die Kombination verschiedener Zugänge zum Forschungsgegenstand meint, besteht Unbestimmtheit darüber, was genau in welchem Maße kombiniert wird.[2] Auch der Begriff Triangulation ist nicht völlig eindeutig, sondern wird von verschiedenen Autoren in variierenden Bedeutungszusammenhängen und mitunter synonym mit dem „integrativen Ansatz" nach Seipel/Rieker (2003) oder dem *mixed methods approach* nach Jick (1979) verwendet. Im Folgenden werden nun zwei theoretische Modelle der Triangulation vorgestellt, um diese dann in einem zweiten Schritt auf die COMBIH-Studie anzuwenden.

Ein Triangulationsmodell stammt von Ivankova/Creswell (2009). Ihr *mixed-methods* Ansatz schreibt nur solchen Forschungsprojekten triangulativen Charakter zu, die qualitativ und quantitativ gewonnene Datensätze miteinander kombinieren. Das *explanatory* und das *exploratory design* legen den Fokus auf nur eine Forschungsmethode (quantitativ oder qualitativ) und betrachten die jeweils andere lediglich als der Hauptstudie zuarbeitende Datenerhebungsmethode. Im *triangulation design* hingegen findet die qualitative und quantitative Datensammlung und -auswertung gleichberechtigt statt. In der Phase der Dateninterpretation werden die beiden Ansätze kombinativ zusammengeführt.

Das zweite Modell findet sich in Flick (2011), der an Denzins Unterteilung in vier Formen der Triangulation von 1970 anknüpft. Als triangulativ wird demnach angesehen,

- wenn mehrere Forscherinnen und Forscher eingesetzt werden, um Daten zu erheben (Forscher-Triangulation),

2 Eine ausführliche Diskussion hierzu findet sich in Flick (2011).

- wenn verschiedene Datensätze kombiniert werden, die an verschiedenen Orten, zu verschiedenen Zeitpunkten und an verschiedenen Personen(-gruppen) gewonnen wurden (Daten-Triangulation),
- wenn sich Forscherinnen und Forscher mit mehreren, forschungsparadigmatisch divergierenden Perspektiven und Hypothesen einem Untersuchungsgegenstand nähern (Theorien-Triangulation)
- und wenn verschiedene methodische Zugänge miteinander kombiniert werden (Methoden-Triangulation), was sowohl innerhalb von als auch zwischen Forschungsparadigmen erfolgen kann (*within-method* bzw. *between-method*).

Die Kombination verschiedener Forschungszugänge verfolgt in der Regel zwei Absichten (Flick 2011: 9). Sie dient entweder der Validierung von Daten bzw. Erhebungsinstrumenten (hiermit soll Konvergenz erzielt werden) oder sie dient der Erkenntniserweiterung (dies führt entweder zu komplementären oder zu divergierenden Ergebnissen). Während erstere Funktion vor allem zu Beginn der Triangulationsforschung als Hauptziel gesehen wurde, fand mittlerweile ein Paradigmenwechsel statt, der nun die erweiterten Erkenntnismöglichkeiten in den Mittelpunkt rückt. Die vorliegende Studie folgt dem Konzept der Triangulation nach Flick, da sich Triangulation als *umbrella term* verstehen lässt, innerhalb dessen sich Ivankovas/Creswells Ansatz zwar im Sinne der *between-method* verorten lässt, der darüber hinaus jedoch vielfältige Formen der Triangulation in Bezug auf Forscher, Datensätze und Theorien bietet.

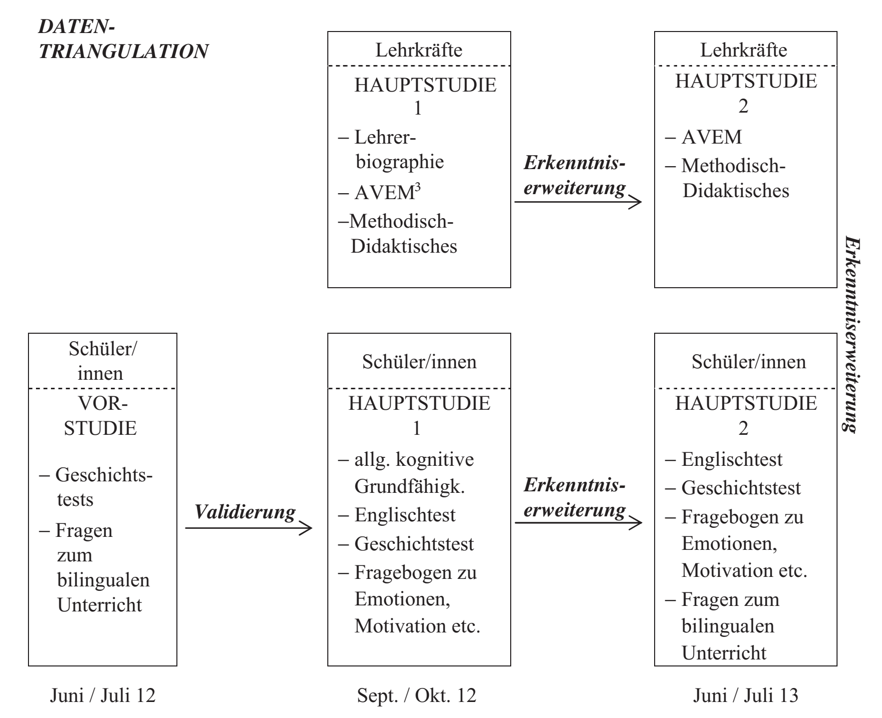
5.2 Triangulation in der COMBIH-Studie

5.2.1 Forscher-Triangulation

Um den bisherigen Forschungsstand zum bilingualen Sachfachunterricht systematisch zu erweitern, wurde die Expertise verschiedener Fachdisziplinen vereinigt. So wurden insbesondere die grundlegenden englisch- und geschichtsdidaktischen Fragestellungen der Studie sowie die Entwicklung der Leistungstests in den Sachfächern von den Fachdidaktikabteilungen Englisch und Geschichte der Pädagogischen Hochschule Ludwigsburg betreut. Für die zentralen psychologischen Konstrukte in den Fragebögen bedurfte es der Expertise der Pädagogischen Psychologie. Dies, sowie die Entwicklung des Studiendesigns, die Betreuung der psychometrischen Verfahren und statistisch-analytische Unterstützung wurde von der Abteilung Empirische Bildungsforschung und Pädagogische Psychologie der Universität Tübingen geleistet. Diese interdisziplinäre Zusammenarbeit im Rahmen des Kooperativen Promotionskollegs „Effektive

Lehr-Lernarrangements“ führte zur Integration verschiedenster Perspektiven und Forschungstraditionen und somit zu einer möglichst angemessenen Annäherung an den Forschungsgegenstand.

Abb. 1: Daten-Triangulation in der COMBIH-Studie.



5.2.2 Daten-Triangulation

Auch der Ansatz der Daten-Triangulation wurde in die COMBIH-Studie integriert. Dieser verfolgte zwei Absichten (vgl. Abb. 1). Erstens diente die Vorstudie der Validierung des neu entwickelten Test- und Befragungsmaterials. Nach der Auswertung der Pilotierungsdaten wurden die Aufgaben des Geschichtstests verbessert sowie eine Selektion der psychometrisch besten, den Gütekriterien empirischer Forschung entsprechenden Items zum bilingualen Unterricht vorgenommen. Diese so optimierten Instrumente wurden dann in die Sammlung

3 Arbeitsbezogene Verhaltens- und Erlebensmuster (vgl. Abschnitt 4.3).

der übrigen (bereits in anderen Studien validierten) Instrumente für die Hauptstudie aufgenommen. Zweitens diente der Ansatz der Daten-Triangulation dazu, verschiedene Perspektiven miteinander zu kombinieren und somit eine Erkenntniserweiterung herbeizuführen. Dies wurde in COMBIH durch die folgenden Maßnahmen erreicht.

Längsschnittliches Design

Schüler- und Lehrerdaten wurden zu zwei Messzeitpunkten erhoben. Der erste Messzeitpunkt dient der Erfassung der Ausgangslage der beiden Schülergruppen zu Beginn des bilingualen Unterrichts, die zweite Messung erlaubt Aussagen über mögliche Veränderungen. Mit dem Geschichtstest wird beispielsweise zunächst das Vorwissen der Schülerinnen und Schüler zu den Inhalten der Klasse 8 überprüft, um dann mithilfe der Messung am Ende des Schuljahres den Leistungszuwachs erfassen zu können. Dasselbe trifft auf die Entwicklung der Englischkompetenzen und der Schülermotivation und Emotionen bezüglich der Fächer Englisch und Geschichte zu. Würden die Schülerinnen und Schüler nur am Ende des Schuljahres, also nach erfolgtem bilingualen Geschichtsunterricht, befragt und getestet, wüsste man zwar, wie die Lerner diese Unterrichtsform erlebt und welche Leistungen sie in Geschichte und Englisch erbracht haben. Es wären jedoch keine Informationen darüber vorhanden, ob diese Unterschiede nicht bereits vor Beginn des bilingualen Geschichtsunterrichts bestanden. Zwei Messzeitpunkte hingegen gestatten es – sofern die entscheidenden konfundierenden Variablen kontrolliert werden – Aussagen über die Effekte des bilingualen Geschichtsunterrichts zu treffen: Wie verändern sich Leistungen, Motivation und Emotionen von bilingual unterrichteten Lernern durch den bilingualen Geschichtsunterricht?

Leistungstests und Fragebogen

Die Schülerinnen und Schüler bearbeiteten zu beiden Messzeitpunkten Leistungstests und einen Fragebogen. Bezüglich der Englischleistungen wurde ein C-Test zur Erfassung der allgemeinen Englischkompetenz eingesetzt sowie ein Hörverstehenstest, da bezüglich dieses Kompetenzbereichs besonders große Unterschiede zwischen bilingual und einsprachig unterrichteten Schülerinnen und Schülern erwartet werden. Das Geschichtswissen wurde mithilfe eines Lückentexts und *Multiple Choice*-Aufgaben erhoben. Der Geschichtstest zu Inhalten aus Klasse 7 soll zeigen, wie gut die Lerner bisher in Geschichte waren. Der Test zu Inhalten aus Klasse 8 dient zum ersten Messzeitpunkt als Vorwissenstest und zu Messzeitpunkt zwei als Maß für den Zuwachs an Geschichtswissen. Um über das Sachwissen hinaus auch Kompetenzen im Fach Geschichte berücksichtigen

zu können, wurden Fragen zur historischen De- und Rekonstruktionskompetenz integriert. Zusätzlich wurden als Indikatoren der Lernvoraussetzungen zwei kognitive Grundfähigkeitstests eingesetzt. Die Tests sind von besonderer Bedeutung, da fundierte, objektive Aussagen zu Schülerleistungen getroffen werden können, die mithilfe der Erhebung von Schulnoten und dem akademischen Selbstkonzept nicht möglich wären. Mit dem Schülerfragebogen wurden mehrere Aspekte erfasst: zum einen die zentralen Konstrukte Motivation und Emotionen in den Fächern Englisch und Geschichte. Zum anderen wurden Fragen integriert, die die Schüler dazu aufforderten, ihren Englisch- und Geschichtsunterricht sowie die Lehrkräfte einzuschätzen. So kann ein Bild davon gewonnen werden, wie Unterricht in den verschiedenen Klassen der Stichprobe ablief.[4] Außerdem dienen diese Fragen der Kontrolle von Schülerleistung, Motivation und Emotionen. Mithilfe des umfangreichen Test- und Befragungsmaterials sollte der Forschungsgegenstand so umfassend und präzise wie möglich erfasst werden.

Schüler- und Lehrerperspektive

In der Studie wurden sowohl Schüler- als auch Lehrerdaten erhoben. Im Lehrerfragebogen befanden sich Fragen zu methodisch-didaktischen Aspekten des (bilingualen) Geschichtsunterrichts, Fragen zur Person, zur Lehrerbiographie, zu arbeitsbezogenen Verhaltens- und Erlebensmustern und eine Einschätzung der an der Studie teilnehmenden Klasse. Mithilfe dieser rund 160 Items sollte der Geschichtsunterricht der Schülerinnen und Schüler so ausgewogen und differenziert wie möglich erfasst werden. Weiterhin diente die Befragung der Lehrkräfte der Kontrolle von Kontextvariablen. Denn Lehrkräfte und die damit in Verbindung stehende Qualität des Unterrichts nehmen Einfluss sowohl auf motivationale und emotionale Aspekte als auch auf das Kompetenzerleben und die Leistungsentwicklung der Schülerinnen und Schüler (Helmke/Schrader 2001, Kunter et al. 2011). So können Aussagen über Motivation, Emotionen und Schülerkompetenzen getroffen werden, die unabhängig sind von mit der Lehrkraft in Verbindung stehenden Faktoren. Denn sollten diese Faktoren systematisch zwischen bilingual und einsprachig unterrichtenden Lehrkräften variieren, wäre dies eine weitere Varianzquelle, die es bei der Isolation der Effekte des bilingualen Unterrichts zu kontrollieren gilt.

4 Einschränkend muss eingeräumt werden, dass lediglich die Wahrnehmung der Schüler (und Lehrkräfte) berichtet werden kann und unklar bleibt, inwieweit diese mit den tatsächlichen unterrichtlichen Vorgängen übereinstimmen.

Bilingual unterrichtete Klassen und zwei Vergleichsgruppen
Um die Effekte des bilingualen Geschichtsunterrichts auf die Entwicklung von Englisch- und Geschichtskompetenzen, Schülermotivation und Emotionen zu bestimmen, bedarf es mindestens einer Vergleichsgruppe deutschsprachig unterrichteter Schülerinnen und Schüler. Würden Ergebnisse auf eine deutlich positivere Entwicklung von Kompetenzen, Motivation und Emotionen bilingual unterrichteter Schülerinnen und Schüler in den Fächern Englisch und Geschichte verweisen, könnte dies das Konzept des bilingualen Lehrens und Lernens weiter stärken und die Ausweitung des bilingualen Unterrichtsangebots unterstützen. Durch die Einbindung von zwei Vergleichsgruppen – Parallelklassen und Klassen aus Schulen ohne bilinguales Angebot – wird bezweckt, mögliche Unterschiede innerhalb der nicht bilingual unterrichteten Schülerschaft zu untersuchen. Diese könnten zu Ungunsten der Parallelklassenschüler ausfallen und eine Konsequenz der positiven Selektion der bilingual unterrichteten Lerner an den bilingualen Zugschulen darstellen.

Gymnasien aus ganz Baden-Württemberg
Die an der Studie beteiligten Gymnasien sind über ganz Baden-Württemberg verteilt und tragen somit zum Ansatz der Daten-Triangulation bei, wonach verschiedene Datensätze an unterschiedlichen Orten gewonnen werden. Die Datenerhebung fand sowohl an Gymnasien statt, die an der hessischen Grenze liegen, als auch an Schulen am Bodensee und in Freiburg; an Gymnasien, die in der Nähe Frankreichs und in der Nähe Bayerns liegen. Darüber hinaus waren an der Studie sowohl städtische als auch ländliche Gymnasien beteiligt, durchschnittlich große Schulen mit rund 1000 Schülerinnen und Schülern und größere, bis zu elfzügige Schulen. Diese unterschiedlichen und recht gleichmäßig über Baden-Württemberg verstreut gelegenen Gymnasien decken ein weites Spektrum der Zielpopulation ab. Allerdings stellen die Schulen keine repräsentativ und zufällig gezogene Stichprobe dar (vgl. Abschnitt 4.1), weshalb die gewonnen Daten nur begrenzt auf die Grundgesamtheit übertragen und verallgemeinert werden können.

5.2.3 Zusammenführung der triangulativen Aspekte
Bei der Analyse der Daten und Interpretation der gewonnenen Ergebnisse von COMBIH werden alle obig berichteten triangulativen Aspekte gleichermaßen berücksichtigt: die unterschiedlichen disziplinären Perspektiven, die beiden Messzeitpunkte, zu denen eine sehr große Stichprobe verteilt über das ganze Bundesland untersucht wurde, die objektiven Leistungstests sowie die standardisiert erfassten Selbstberichte zum Unterrichtserleben, zur Motivation und zu

Emotionen, die Schüler- und Lehrerperspektive sowie die unterschiedlichen Vergleichsgruppen. In diesem Sinne kann in Anlehnung an die Arbeiten von Ivankova/Creswell die Gewichtung der verschiedenen Datenquellen bei COMBIH – „the relative importance or priority given to each type of data“ (2009: 138) – als gleichgewichtet bezeichnet werden. Somit wird einer wichtigen Forderung dieses forschungsparadigmatischen Ansatzes Rechnung getragen, wonach der Mehrwert der Triangulation in der bedeutsamen Integration und nicht in einer bloßen Akkumulation von verschiedenen Zugängen zum Forschungsgegenstand liegt.

Literatur

Abendroth-Timmer, Dagmar (2004). Evaluation bilingualer Module aus Schülerperspektive: zur Lernbewusstheit und ihrer motivationalen Wirkung. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 9/2: 1-27. Online: http://zif.spz.tu-darmstadt.de/jg-09-2/beitrag/Abendroth2.htm [10.3.2014].

Alonso, Esmeralda / Grisaleña, Jesús / Campo, Alejandro (2008). Plurilingual Education in Secondary Schools: Analysis of Results. *International CLIL Research Journal* 1/1: 36-49. Online: http://www.icrj.eu/11/article3.html [10.3.2014].

Badertscher, Hans / Bieri, Thomas (2009). *Wissenserwerb im Content and Language Integrated Learning. Empirische Befunde und Interpretationen*. Bern: Haupt.

Baumert, Jürgen / Trautwein, Ulrich / Artelt, Cordula (2003). Schulumwelten. Institutionelle Bedingungen des Lehrens und Lernens. In: Baumert, Jürgen et al. (Hrsg.). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Opladen: Leske & Budrich, 259-330.

Bertram, Christiane (2012). Wirksamkeit von Zeitzeugenbefragungen im Geschichtsunterricht – ein Beitrag für die empirische Geschichtsdidaktik. *Geschichte für heute. Zeitschrift für historisch-politische Bildung* 5/3, 21-33.

Bonnet, Andreas (2012). Towards an Evidence Base for CLIL. How to Integrate Qualitative and Quantitative as well as Process, Product and Participant Perspectives in CLIL Research. *International CLIL Research Journal* 4/1: 66-78. Online: http://www.icrj.eu/14/article7.html [10.3.2014].

Breidbach, Stephan / Viebrock, Britta (2012). CLIL in Germany: Results from Recent Research in a Contested Field of Education. *International CLIL Research Journal* 1/4: 5-16. Online: http://www.icrj.eu/14/article1.html [10.3.2014].

Bruton, Anthony (2011). Is CLIL so Beneficial, or Just Selective? Re-evaluating Some of the Research. *System* 39/4, 523-532.

Costa, Francesca / D'Angelo, Lauretta (2011). CLIL: A Suit for All Seasons. *Latin American Journal of Content & Language* 4/1, 1-13.

Coyle, Do / Hood, Philip / Marsh, David C. (2010). *CLIL. Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Dallinger, Sara (2013). Die Einstellung der Eltern zum bilingualen Sachfachunterricht. Eine quantitative Studie an Realschulen in Baden-Württemberg. In: Hollm, Jan /Hüttermann, Armin / Keßler, Jörg-U. / Schlemminger, Gérald / Ade-Thurow, Benjamin (Hrsg.). *Bilinguales Lehren und Lernen in der Sekundarstufe I. Sprache Sachfach und Schulorganisation*. Landau: Verlag Empirische Pädagogik, 268-286.

DESI-Konsortium (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International (DESI)*. Online: http://www.dipf.de/de/projekte/pdf/biqua/desi-zentrale-befunde [10.3.2014].

Fehling, Sylvia (2008). *Language Awareness und bilingualer Unterricht. Eine komparative Studie*. Frankfurt/Main: Lang.

Flick, Uwe (2011). *Triangulation. Eine Einführung*. 3. Aufl. Wiesbaden: VS Verlag.

Gläser-Zikuda, Michaela / Seidel, Tina / Rohlfs, Carsten / Gröschner, Alexander / Ziegelbauer, Sascha(Hrsg.) (2012). *Mixed Methods in der empirischen Bildungsforschung*. Münster: Waxmann.

Götz, Thomas / Cronjäger, Hanna / Frenzel, Anne C. / Lüdtke, Oliver / Hall, Nathan C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology* 35, 44-58.

Götz, Thomas / Frenzel, Anne C. (2010). Über- und Unterforderungslangeweile im Mathematikunterricht. *Zeitschrift Empirische Pädagogik* 24/2, 113-134.

Heller, Kurt A. / Perleth, Christoph (2000). *Kognitiver Fähigkeitstest für 4. – 12. Klassen. Revision (KFT 4-12+R)*. Göttingen: Hogrefe.

Helmke, Andreas / Schrader, Friedrich-Wilhelm (2001). Determinanten der Schulleistung. In: Rost, Detlef H. (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. Weinheim: Beltz PVU, 81-90.

Hollm, Jan / Hüttermann, Armin / Keßler, Jörg-U. / Schlemminger, Gérald (2010). *Zwischenbericht der wissenschaftlichen Begleitung zum Schulversuch „Bilinguale Züge an Realschulen" in Baden-Württemberg. Projekttitel: BiliReal 2012*. Online: http://www.ph-ludwigsburg.de/fileadmin/subsites/2b-engl-t-01/Bili/Zwischenbericht_Schulversuch_BLLZuege_RS_Ba-Wue.pdf [10.3.2014].

Hollm, Jan / Hüttermann, Armin / Keßler, Jörg-U. / Schlemminger, Gérald / Ade-Thurow, Benjamin (Hrsg.) (2013). *Bilinguales Lehren und Lernen in der*

Sekundarstufe I. Sprache Sachfach und Schulorganisation. Landau: Verlag Empirische Pädagogik.

Ivankova, Nataliya V. / Creswell, John W. (2009). Mixed Methods. In: Heigham, Juanita / Croker, Robert (Hrsg.). *Qualitative Research in Applied Linguistics*. Basingstoke: Palgrave Macmillan, 135-161.

Jick, Todd D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly* 24/4, 602-611.

Kollenrott, Anne I. (2008). *Sichtweisen auf deutsch-englisch bilingualen Geschichtsunterricht. Eine empirische Studie mit Fokus auf interkulturelles Lernen*. Frankfurt/Main: Lang.

Köller, Olaf / Leucht, Michael / Pant, Hans A. (2012). Effekte bilingualen Unterrichts auf die Englischleistungen in der Sekundarstufe I. *Unterrichtswissenschaft* 40/4, 334-350.

Krapp, Andreas (2002). Structural and Dynamic Aspects of Interest Development: Theoretical Considerations from an Ontogenetic Perspective. *Learning and Instruction* 12, 383-409.

Kunter, Mareike / Baumert, Jürgen / Blum, Werner / Klusmann, Uta / Krauss, Stefan / Neubrand, Michael (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

Lamsfuß-Schenk, Stefanie (2008). *Fremdverstehen im bilingualen Geschichtsunterricht. Eine Fallstudie*. Frankfurt/Main: Lang.

Lasagabaster, David / Sierra, Juan M. (2009). Language Attitudes in CLIL and Traditional EFL Classes. *International CLIL Research Journal* 2/1, 4-17. Online: http://www.icrj.eu/12/article1.html [10.3.2014].

Nold, Günter / Hartig, Johannes / Hinz, Silke / Rossa, Henning (2008). Klassen mit bilingualem Sachfachunterricht: Englisch als Arbeitssprache. In: DESI-Konsortium (Hrsg.). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz, 451-457.

Pekrun, Reinhard (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review* 18, 315-341.

Pérez-Cañado, María L. (2012). CLIL research in Europe: past, present, and future. *International Journal of Bilingual Education and Bilingualism* 15/3, 315-341.

Raudenbush, Stephen W. / Bryk, Anthony S. (2002). *Hierarchical Linear Models*. Thousand Oaks: Sage.

Riemer, Claudia (2011). Empirie und Fremdsprachenforschung: Herausforderung Forschungsmethodik. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Krumm, Hans-Jürgen (Hrsg.). *Fremdsprachen lehren und lernen: Rück- und Ausblick. Arbeitspapiere der 30. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 194-203.

Rumlich, Dominik (2013a). Students' General English Proficiency prior to CLIL: Empirical Evidence for Substantial Differences between Prospective CLIL and Non-CLIL Students in Germany. In: Breidbach, Stephan / Viebrock, Britta (Hrsg.). *Content and Language Integrated Learning (CLIL) in Europe. Research Perspectives on Policy and Practice.* Frankfurt/Main: Lang, 181-201.

Rumlich, Dominik (2013b). *The Development of CLIL and Non-CLIL Students' General Language Proficiency: Findings from the Large-scale Longitudinal Study DENOCS*. Vortrag auf dem Kongress der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) in Augsburg am 27.09.2013.

Rumlich, Dominik (2014). Prospective CLIL and non-CLIL Students' Interest in English (Classes): A Quasi-experimental Study on German Sixth-graders. In: Breeze, Ruth / Martínez Pasamar, Concepción / Llamas Saíz, Carmen / Tabernero Sala, Cristina (Hrsg.). *Integration of Theory and Practice in CLIL*. Amsterdam: Rodopi, 75-95

Schaarschmidt, Uwe / Fischer, Andreas W. (2008). *Arbeitsbezogenes Verhaltens- und Erlebensmuster (AVEM)*. Frankfurt/Main: Pearson.

Schreiber, Waltraud / Körber, Andreas / Borries van, Bodo / Krammer, Reinhard / Leutner-Ramme, Sibylla / Mebus, Sylvia / Schöner, Alexander / Ziegler, Béatrice(2006). *Historisches Denken. Ein Kompetenz-Strukturmodell*. Neuried: ars una.

Seipel, Christian / Rieker, Peter (2003). *Integrative Sozialforschung. Konzepte und Methoden der qualitativen und quantitativen empirischen Forschung*. Weinheim: Juventa.

Snijders, Tom / Bosker, Roel (2011). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles: Sage.

Wigfield, Allan L. / Eccles, Jacquelynne S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology* 25, 68-81.

Wild, Elke / Hofer, Manfred (2000). Elterliche Erziehung und die Veränderung motivationaler Orientierungen in der gymnasialen Oberstufe und der Berufsschule. In: Schiefele, Ulrich / Wild, Klaus-Peter (Hrsg.). *Interesse und Lernmotivation*. Münster: Waxmann, 31-52.

Witzigmann, Stéfanie (2011). *Bildende Kunst in der Zielsprache Französisch als Einstieg ins bilinguale Lehren und Lernen. Eine explorative Studie in einer fünften Realschulklasse*. Frankfurt/Main: Lang.

Zydatiß, Wolfgang (2007). *Deutsch-Englische Züge in Berlin (DEZIBEL). Eine Evaluation des bilingualen Sachfachunterrichts an Gymnasien*. Frankfurt/Main: Lang.

Zydatiß, Wolfgang (2012). Linguistic Thresholds in the CLIL Classroom? The Threshold Hypothesis Revisited. *International CLIL Research Journal* 4/1, 17-28. Online: http://www.icrj.eu/ 14/article2.html [10.3.2014].

Triangulation: Möglichkeiten, Grenzen, Desiderate

Karin Aguado

Die Triangulation erfreut sich in jüngster Zeit insbesondere beim wissenschaftlichen Nachwuchs einer stetig wachsenden Popularität, was angesichts der Komplexität der Erforschung von fremdsprachenbezogenen Lehr- und Lernprozessen grundsätzlich zu begrüßen ist. Da diese Forschungsstrategie jedoch mit einer erheblichen Mehrarbeit verbunden ist, gilt zunächst grundsätzlich abzuwägen, ob ihre Anwendung der eigenen Fragestellung tatsächlich angemessen ist. Im vorliegenden Beitrag werden zunächst ausgewählte theoretische und praktische Aspekte der Triangulation skizziert. Da in der qualitativen Forschung ja der Gegenstand bzw. das Erkenntnisinteresse eine entscheidende Rolle bei der Bestimmung der für die empirische Untersuchung einzusetzenden methodischen Verfahren spielt, wird im Anschluss die ‚Angemessenheit' als zentrales Gütekriterium empirischer Forschung diskutiert. Diese Überlegungen sollen bei der Entscheidung für oder gegen die Anwendung der Triangulationsstrategie helfen. Abschließend werden einige offene Fragen und Desiderate formuliert, deren weitere Diskussion meiner Ansicht nach lohnt.

1. Einleitung

Triangulation hat sich – zumindest in primär qualitativ-interpretativ angelegten empirischen Studien – inzwischen zu einer weit verbreiteten Forschungsstrategie entwickelt und zwar nicht zuletzt, weil das Verstehen von komplexen Zusammenhängen im Sinne eines methodisch kontrollierten Fremdverstehens ein zentrales Erkenntnisprinzip qualitativer Forschung darstellt (vgl. dazu Arbeitsgruppe Bielefelder Soziologen 1976). Grundvoraussetzung für die Akzeptanz der Triangulation sowohl auf Seiten derjenigen, die sie anwenden, als auch auf Seiten derjenigen, welche die mit ihrer Hilfe erzielten Ergebnisse rezipieren, ist eine methodologische Offenheit, in der Ausschließlichkeits- oder Prioritätenansprüche einer Methodologie respektive einer Methode gegenüber einer anderen keine Rolle spielen (vgl. dazu auch Lamnek 2005: 279). Insbesondere in Forschungen, die faktorenkomplexe Gegenstände wie z.B. das Lehren und Lernen von Sprachen untersuchen, scheinen mehrmethodische Vorgehensweisen angemessen zu sein, weil sie eine höhere Adäquanz ermöglichen und somit auch für alltagsweltliche Akteure verstehbar und vernünftig erscheinen (vgl. Schütz 1974). Ob und inwieweit die mit triangulativen Verfahren verbundenen

Erwartungen berechtigt und erfüllbar sind, wird im vorliegenden Sammelband anhand ausgewählter empirischer Studien diskutiert.

Es besteht das Problem, dass die Triangulation von nicht wenigen Kollegen und Kolleginnen inzwischen nicht nur als *Merkmal*, sondern sogar als *Gütekriterium* qualitativer Forschung betrachtet wird (vgl. z.B. Mayring 2002: 147). Ich halte dies u.a. deshalb für problematisch, weil dadurch insbesondere wenig erfahrene Forscherinnen und Forscher[1] den Eindruck gewinnen (müssen), dass ein mehrmethodisches Vorgehen für eine hochwertige, aktuellen forschungsmethodologischen Entwicklungen verpflichtete qualitative Forschung zwingend erforderlich sei – schließlich will man ja als qualitativ arbeitende Forscherin bzw. als qualitativ arbeitender Forscher die gültigen Gütekriterien einhalten und gute Forschung machen. Es soll jedoch im Folgenden gezeigt werden, dass Triangulation weder ein Wert an sich noch ein Gütekriterium qualitativer Forschung ist!

Da die wesentlichen Formen, Merkmale und Funktionen der Triangulationsstrategie bereits in der Einleitung der beiden Herausgeberinnen Daniela Elsner und Britta Viebrock als auch in dem Beitrag von Julia Settinieri (Kap. 2 in diesem Band) ausführlich dargestellt worden sind, möchte ich sie im Folgenden nur kurz skizzieren (vgl. auch Aguado/Riemer 2001 oder Aguado 2009). Stattdessen soll die Aufmerksamkeit in erster Linie auf terminologische und konzeptionelle sowie auf praktische Überlegungen hinsichtlich der Möglichkeiten und Grenzen gerichtet werden, die in der einschlägigen Literatur im Zusammenhang mit der Triangulationsstrategie diskutiert werden.

In Bezug auf die Terminologie stellt sich zunächst einmal die Frage nach der Angemessenheit der Übertragung des Ausdrucks ‚Triangulation' aus der Geodäsie auf die Forschungsmethodologie der Sozialwissenschaft, der Erziehungswissenschaft oder der Fremdsprachendidaktik – also allesamt Disziplinen, die vollkommen andere Gegenstände, Erkenntnisinteressen und Fragestellungen haben als die in erster Linie an exakten Messungen interessierten Naturwissenschaften. Hinzu kommt das Problem, dass die verschiedenen Termini ‚Triangulation', *mixed methods*, Mehr-Methoden-Design oder ‚Methodenintegration' – auch von einschlägigen Kolleginnen und Kollegen wie z.B. Denzin/Lincoln (1994) – vielfach synonym gebraucht werden, was jedoch ungenau ist und daher vermieden werden sollte, und zwar nicht zuletzt weil z.B. Studierende oder Nachwuchswissenschaftler/innen, welche die historische Entwicklung der hier

1 In den Beiträgen dieses Bandes werden maskuline und feminine Formen in unterschiedlicher Weise verwendet. Unter der Voraussetzung, dass jeweils beide Geschlechter gemeint sind, haben die Herausgeberinnen entschieden, diese Vielfalt bestehen zu lassen.

relevanten Konzepte, Konstrukte und Strategien nicht kennen, eine verkürzte und damit z.T. unzutreffende Vorstellung davon erhalten.

Der Zusammenhang zwischen Gegenstand und Methode ist komplex und vielfältig und kann hier nicht im Detail diskutiert werden. Zentral für den vorliegenden Kontext ist die Tatsache, dass in der qualitativen Forschung der Gegenstand und nicht die Methode im Zentrum steht. In einem nächsten Schritt soll daher das Kriterium der Angemessenheit diskutiert werden, das in der qualitativen Forschung grundsätzlich für die Auswahl *jeder* Methode der Erhebung, Aufbereitung und Auswertung von Daten zu beachten ist und das daher selbstverständlich auch für die Auswahl und Kombination mehrerer Methoden innerhalb einer Studie gilt.

Abschließend sollen noch einmal der in der Diskussion bisher erreichte Minimalkonsens hinsichtlich der Triangulationsstrategie und Möglichkeiten sowie Erfordernisse ihrer Weiterentwicklung angesprochen werden (für eine aktuelle Darstellung siehe Kuckartz 2014). In diesem Zusammenhang werden auch einige der im forschungsmethodologischen Diskurs bereits mehrfach formulierten Desiderate aufgegriffen.

2. Terminologisches, Konzeptionelles, Praktisches

Zu dem Terminus ‚Triangulation', der sich inzwischen etabliert hat (was auch an dem Titel des vorliegenden Sammelbands und der ihr zugrunde liegenden DGFF-Sommerschule erkennbar ist), ist zu sagen, dass die Übertragung aus einer Naturwissenschaft auf die Sozialwissenschaften sowie auf andere, vorwiegend qualitativ arbeitende Disziplinen streng genommen unangemessen ist (vgl. dazu auch Blaikie 1991). Der Grund ist, dass es bei der Erdvermessung um eine formal berechenbare, durch die Einnahme mehrerer Perspektiven möglichst genaue Positionsbestimmung geht, während in sozial- oder lehr-/lernwissenschaftlichen Forschungen auf eine inhaltlich beschreibende, verstehende und/oder interpretative Untersuchung des jeweiligen Gegenstands abgezielt wird. D.h. ursprünglich, also in den Naturwissenschaften und in anderen bevorzugt analytisch-nomologisch arbeitenden Disziplinen, bestand die Hauptfunktion der Triangulationsstrategie in der Vermeidung von Verzerrungen bzw. Messfehlern und somit im Ausschluss von Messartefakten (vgl. Campbell/Fiske 1959). Daraus resultiert auch die eher defizitbasierte Annahme, dass durch die Anwendung mehrerer Methoden die jeder Einzelmethode inhärenten Schwächen kompensiert werden könnten – eine Sichtweise auf Methoden, die der qualitativen Forschung eher fremd ist oder sein sollte, wählt man eine Methode doch wegen ihrer Eignung und ihrer Vorzüge aus und nicht damit sie Unzulänglichkeiten

ausgleicht. Bereits an dieser Stelle werden grundlegende Unterschiede zwischen quantitativen und qualitativen Sichtweisen auf die möglichen Funktionen der Triangulation – also Validierung vs. Erkenntniserweiterung – deutlich. Fielding und Fielding (1986) zweifelten bereits Mitte der 1980er Jahre daran, dass man die mit einer Methode gewonnenen Daten durch die mit einer anderen Methode gewonnenen Daten korrigieren könnte. Wenn man sich überhaupt auf eine solche kompensatorische Argumentation einlassen will, erscheint es eher angemessen anzunehmen, dass eine mehrmethodische Herangehensweise der Kompensation von blinden Flecken, aber nicht der von Schwächen dienen kann. Dass ein Gegenstand durch die zu seiner Untersuchung eingesetzten Methoden konstituiert wird, gehört in der forschungsmethodologischen Diskussion inzwischen zum Konsens: Jede Methode ermöglicht die Betrachtung eines spezifischen Ausschnitts, einer bestimmten Perspektive, einer anderen Facette des jeweiligen Gegenstands und bestimmt somit, was wir über ihn erfahren. Insbesondere auch deshalb erscheint die Kombination mehrerer Methoden sinnvoll. Die Anerkennung dieser Tatsache bringt eine Reihe von Konsequenzen für die empirische Forschung mit sich, die meiner Ansicht nach bisher noch nicht ausreichend diskutiert worden sind. So ergibt sich für die Methoden-Kombination innerhalb einer Studie die grundsätzliche Frage, ob es überhaupt möglich ist, mittels verschiedener Methoden *genau denselben* Gegenstand zu untersuchen. Müsste es – unter der zuvor genannten Prämisse, dass ja die Methode den Gegenstand konstituiert – nicht vielmehr so sein, dass verschiedene Methoden zwingend verschiedene Gegenstände untersuchen? Hier könnte man entgegnen, dass es natürlich auf den Gegenstand ankommt, d.h. es gibt einfache und komplexe Gegenstände, und wie ein Gegenstand definiert wird, was dazu gehört bzw. aus welchen Komponenten er sich zusammensetzt, entscheidet letztlich der/die Forschende selbst bzw. die *scientific community*, der er/sie angehört. Der Gegenstand ‚Lehren und Lernen von Fremdsprachen' ist beispielsweise außerordentlich komplex (vgl. zur Faktorenkomplexion in der Fremdsprachenforschung Grotjahn 2006: 248), und um dieser Komplexität angemessen Rechnung zu tragen, bedarf es differenzierter forschungsmethodologischer Überlegungen und Entscheidungen sowie methodisch geeigneter Instrumente und Verfahren, die in der Regel individuell und spezifisch für den zu untersuchenden Gegenstand entwickelt werden, was jedoch nicht unerhebliche Probleme hinsichtlich der Replikation mit sich bringt (vgl. auch Abschnitt 3).

Der Hintergrund der Frage, ob mittels verschiedener Methoden exakt derselbe Gegenstand oder verschiedene Facetten eines Gegenstands erfasst werden, ist der, dass von ihrer Beantwortung abhängt, welche Funktion der Mehrmethodeneinsatz hat: Geht es um die kumulative Validierung oder um die umfassende

Erfassung, Beschreibung und Erklärung (vgl. zur Begrifflichkeit Kelle/Erzberger 2000)?

Ziel der qualitativen Forschung sind sogenannte ‚dichte' Beschreibungen (Geertz 1983), die einen Gegenstand nicht nur in der Tiefe und in ihrem Kern erfassen, sondern die darüber hinaus ihren Rezipientinnen und Rezipienten – sowohl Kolleg/innen als auch Anwender/innen – eine Einschätzung hinsichtlich der Glaubwürdigkeit und der Übertragbarkeit der gewonnenen Erkenntnisse ermöglichen sollen. In diesem Zusammenhang ist offenkundig, dass die Einnahme lediglich einer Perspektive zwar eine sehr detaillierte Betrachtung ermöglicht, aber im Hinblick auf den Gesamtzusammenhang nur beschränkt Rückschlüsse erlaubt, so dass die Einbeziehung zusätzlicher Daten sinnvoll oder gar unverzichtbar ist.

Die bekannteste und in praktisch jeder – und somit auch in der vorliegenden – Abhandlung zum Thema Triangulation zitierte Typologie stammt von Denzin (1970), der vor nunmehr über 40 Jahren erstmals eine Strukturierung dieses methodologischen Konstrukts vornahm und zwischen der Daten-, Forscher-, Theorien- und Methodentriangulation unterschied (vgl. dazu Settinieri in diesem Band; vgl. auch Aguado 2014). Flick (2004) fügte später noch die Perspektiventriangulation hinzu, der jedoch nur ein bedingter Mehrwert zukommt, da allen zuvor genannten Typen die Betrachtung des Gegenstands aus mehreren Perspektiven inhärent ist. Dass im Rahmen einer empirischen Studie alle Typen berücksichtigt werden, kommt grundsätzlich eher selten vor. Dallinger und Jonkmann (in diesem Band) bieten allerdings ein gutes Beispiel für die Integration von drei Typen, nämlich Forscher-, Daten- und Methoden-Triangulation.

Ein seit langem weithin bekanntes Problem der Denzin'schen Klassifikation ist ihre mangelnde definitorische Präzision sowie die fehlende Trennschärfe zwischen den verschiedenen Typen (vgl. dazu auch Birkner/Hackfort 2006). So ist zu beachten, dass grundsätzlich jede Methodentriangulation immer auch zugleich eine Datentriangulation ist, da mit verschiedenen Methoden zwingend auch verschiedene Daten erhoben werden. Von Datentriangulation spricht Denzin jedoch nur dann, wenn die Daten mit derselben Methode erhoben werden, also entweder zu verschiedenen Zeitpunkten oder von verschiedenen Probanden und Probandinnen. Problematisch an dieser Position ist, dass damit praktisch jede Longitudinalstudie, die dieselben Forschungsteilnehmer/innen zu verschiedenen Zeitpunkten (zu Beginn, in der Mitte, am Ende der Studie) beobachtet oder befragt automatisch als Triangulationsstudie bezeichnet werden müsste, was jedoch an der Kernidee der Triangulation vorbeigeht, weil die als zentral benannten Funktionen – Validierung und Erkenntniserweiterung – so nicht erfüllt werden. Stattdessen dienen longitudinale Studien der Erfassung, Dokumentation und im Idealfall auch der Erklärung von Prozessen und Entwicklungen, und dafür sind

nun einmal Daten vonnöten, die zu verschiedenen Zeitpunkte erhoben und zueinander in Beziehung gesetzt, aber nicht miteinander trianguliert werden.

Denzin zufolge sollen verschiedene Verfahren *die eine* richtige oder gültige Sicht auf den Gegenstand ermöglichen. Diese Position wurde u.a. von Silverman (1985) massiv kritisiert, der nicht nur gegen das Konstrukt einer ‚Master-Wirklichkeit' argumentiert, sondern auch gegen die hier zugrundeliegende Annahme, dass überhaupt ein Vollbild eines Gegenstands möglich sei. Hier stellt sich die Frage, wann ein Gegenstand vollständig erfasst ist bzw. wer dies entscheidet. In der einschlägigen forschungsmethodologischen Diskussion ist die Auffassung konsensfähig, dass es keine objektive Realität gibt, sondern dass diese individuell konstruiert bzw. interaktiv ko-konstruiert wird. Eine Aufgabe der qualitativen Forschung ist a) die Re-Konstruktion und b) die Interpretation dieser Konstruktionen. Trotz systematischem und methodisch kontrolliertem Vorgehen und dem Einsatz modernster Mitschnitttechnik bleiben Beobachtungen und Interpretationen durch Forschende ein Stück weit subjektiv. Die Glaubwürdigkeit empirischer Forschung steht daher in einem engen Zusammenhang mit der Anerkennung der Unvermeidbarkeit der Subjektivität und dem gezielten Versuch, reflektiert und angemessen mit dieser Problematik umzugehen und durch eine maximal transparente Dokumentation der Vorgehensweise Nachvollziehbarkeit und damit Intersubjektivität herzustellen. Der erkenntnistheoretische und -praktische Nutzen der Triangulationsstrategie ist hier offenkundig. Von Bedeutung ist auch die Frage nach dem Verhältnis und der Wechselwirkung der verschiedenen Daten, Forscher, Theorien oder Methoden: Handelt es sich um eine bloße Akkumulation oder findet eine bedeutsame Integration statt? Generell wird in der einschlägigen Literatur gegen einen rein technokratischen Umgang mit Methoden argumentiert und die Position vertreten, dass die schlichte Aggregation von Informationen über einen Gegenstand nicht zwingend einen Mehrwert hat oder zu verbesserten Erkenntnissen führt.

Als ein willkommener Nebeneffekt der Triangulation wird die daraus resultierende Demokratisierung von Forschung betrachtet, und zwar insofern als durch die angestrebte Gleichwertigkeit von Perspektiven – z.B. von Forschenden und Forschungsteilnehmern oder von Vertreterinnen und Vertretern unterschiedlicher theoretischer und/oder methodisch-methodologischer Ansätze – Hierarchien aufgebrochen oder gar aufgehoben werden. Auch das Postulat der Gleichwertigkeit methodischen Ansätze und der Methoden als Voraussetzung für eine echte Triangulation wird in der einschlägigen Diskussion immer wieder betont, wobei ich den Aspekt der Parität von Methoden für weniger relevant halte als den Aspekt der Angemessenheit (mehr dazu in Abschnitt 3). Gleichzeitig ist festzustellen, dass die bisher entwickelten und von Mayring (2001) zusammenfassend dargestellten Modelle – nämlich das Vorstudien-, das

Verallgemeinerungs-, das Vertiefungs- und das Triangulationsmodell (für eine ausführliche Darstellung der verschiedenen Modelle siehe Settinieri in diesem Band) – zwar alle mehrmethodisch, aber bis auf das gleichzeitig-synergierende Triangulationsmodell nicht alle triangulierend sind. In dem Vorstudienmodell, das auch als Phasenmodell der Methodenintegration bezeichnet wird, werden zunächst mittels qualitativer Verfahren Hypothesen generiert, die dann mittels quantitativer Verfahren getestet werden. Anders als in der einschlägigen Literatur immer wieder moniert, bedeutet die vorgelagerte Generierung von Hypothesen aus meiner Sicht keineswegs, dass qualitativen Verfahren lediglich Hilfs- oder Dienstleistungsfunktionen zukommen. Im Gegenteil ist die systematische und methodisch kontrollierte Gewinnung von Hypothesen ein unverzichtbarer inhaltlicher und methodischer Schritt, der die gesamte weitere Forschung bestimmt. Die zuvor skizzierte Wahrnehmung seitens der qualitativ Forschenden scheint mir auf einem vollkommen zu Unrecht bestehenden Minderwertigkeitskomplex gegenüber quantitativen Ansätzen zu beruhen.

Grundsätzlich lautet die angesichts der aktuell konstruktivistisch geprägten Sozialforschung vorherrschende Ansicht, dass die Realität nicht nur durch uns, sondern auch durch unsere Methoden konstituiert wird und dass kongruierende Ergebnisse somit eher unwahrscheinlich sind. Erkenntnisse sind also methodeninduziert oder zumindest methodenabhängig, und somit ist die Wahrscheinlichkeit, durch den Einsatz von mehreren Methoden Divergenzen zu erhalten, deutlich größer als die Aussicht auf Konvergenzen. Zusammenfassend lässt sich daher zumindest für die qualitative Forschung konstatieren, dass seit den 1980er Jahren das Ziel der Validierung mittels der Triangulationsstrategie immer seltener verfolgt wird. Für den Fall, dass Validierung und damit Kongruenz angestrebt wird, ist allerdings zu beachten, dass sich unabhängig voneinander zwei oder mehrere Personen irren können. So warnt z.B. Lamnek: „Auch zwei übereinstimmende Befunde erhalten durch die Übereinstimmung keine höheren Wahrheitsgehalt; sie können gleichwohl beide falsch sein.“ Andererseits ist die Wahrscheinlichkeit, „dass zwei gleichlautende Erkenntnisse auf der Basis unterschiedlicher Methoden gleichsinnig falsch sind, […] aufgrund der wahrscheinlichkeitstheoretisch geforderten multiplikativen Verknüpfung sehr gering, wenngleich ein solches Ereignis selbst nicht auszuschließen ist“ (Lamnek 2005: 285).

Neben den zuvor genannten Einschränkungen auf theoretischer, konzeptioneller und funktionaler Ebene gibt es aber auch eine Reihe ganz praktischer Schwierigkeiten mit der Anwendung der mehrmethodischen Triangulationsstrategie. So erfordert sie zunächst einmal noch mehr Methodenkompetenz als die Anwendung einer Einzelmethode. Vor allem die Tatsache, dass die wissenschaftstheoretischen Grundlagen einer jeden Methode gekannt werden müssen,

stellt Forschende, die in anderen wissenschaftlichen ‚Schulen' oder Paradigmen sozialisiert worden sind und somit bestimmte methodisch-methodologische Präferenzen haben, vor große Herausforderungen. Es reicht also nicht, lediglich eine offene Einstellung gegenüber der Kombination oder Integration von mehreren methodischen Verfahren zu haben. Damit mehrmethodische Forschung tatsächlich einen echten Mehrwert haben kann, bedarf es einer entsprechenden Multi-Kompetenz. Dazu gehört die Fähigkeit, begründete Entscheidungen hinsichtlich der Kombinierbarkeit von Methoden treffen zu können und zu wissen, welche Verfahren in welcher Weise miteinander angewendet werden können. Eine nicht ausreichend reflektierte Kombination kann zu Fehlern bei der Anwendung führen und falsche Ergebnisse und damit u.U. fehlerhafte Theoriebildungen zur Folge haben. Hinzu kommt – wie Stefer (in diesem Band: 132) zutreffend feststellt – die Tatsache, dass Forschende „nicht nur in beiden Analysearten bewandert sein müssen, sondern auch die durch die Triangulation entstehende Komplexität zu bewältigen haben. Schließlich ist nicht alles, was sich transformieren, mixen, triangulieren oder integrieren lässt, auch wirklich sinnvoll". So sind bei der Kombination von Methoden verschiedener Paradigmen (*between-method*) insbesondere auch deren jeweilige Eigengesetzlichkeiten sowie die jeweiligen Gütekriterien zu beachten. Beispielsweise unterscheiden sich die *Sampling*-Strategien – d.h. die Art und Weise der Rekrutierung und Auswahl geeigneter Probanden bzw. Forschungsteilnehmerinnen – von quantitativen und qualitativen Verfahren erheblich voneinander, so dass nicht ohne Weiteres von der Validität von Daten auszugehen ist, die qualitativ erhoben und quantitativ ausgewertet worden sind. Die mit der Triangulationsstrategie verbundene Hoffnung auf einen Mehrwert an Qualität, Verlässlichkeit und Glaubwürdigkeit kann nur dann erreicht werden, wenn die Spezifika der jeweiligen Kombinationen angemessen bedacht werden.

Zusammenfassend und mit Bezug auf die Herkunft des Terminus Triangulation sollte grundsätzlich eine eher enge bzw. strenge Definition vertreten werden, der zufolge sich die Strategie auf denselben Gegenstand beziehen muss. Im Unterschied zu Ivankova und Creswell (2009), die nur dann von Triangulation sprechen, wenn quantitative und qualitative Verfahren in einer Studie miteinander kombiniert werden würden, ist es aus meiner Sicht vollkommen irrelevant, ob es sich um *within-method* oder *between- method* Vorgehensweisen handelt.

3. Angemessenheit – das Hauptkriterium für die Anwendung der Triangulationsstrategie

Das im Folgenden diskutierte Gütekriterium der Angemessenheit bezieht sich auf den gesamten Forschungsprozess, d.h. von der Erhebung über die

Aufbereitung bis hin zur Auswertung und zur Interpretation von Daten. Diesem Zentralbegriff qualitativer Forschungen (vgl. auch Brüsemeister 2008) kommt in Bezug auf die Triangulationsstrategie meiner Auffassung nach eine besondere Gültigkeit zu.

Wie in so vielen Bereichen trifft auch in der empirischen Forschung die Losung „Viel hilft viel!“ nur begrenzt zu. So ist der Einsatz möglichst vieler Methoden im Rahmen einer einzigen Studie nicht *per se* ein Wert (vgl. dazu z.B. Lamnek 2005: 291). Vielmehr muss die Anwendung mehrerer Methoden – genau wie die Anwendung einer einzelnen Methode – dem Gegenstand angemessen sein, sich für die Beantwortung der Forschungsfrage eignen und dem übergeordneten Erkenntnisinteresse dienen. Die Entscheidung für die jeweilige methodische Auswahl muss immer entsprechend nachvollziehbar begründet werden – und zwar unabhängig von der Menge der Methoden. Es geht hier um die optimale Passung zwischen dem spezifischen Forschungsgegenstand und der Methode. Dabei ist grundsätzlich zu beachten, dass es selbstverständlich eine Vielzahl von Fragestellungen gibt, für die die Anwendung einer Einzelmethode nicht nur ausreichend, sondern auch besser für ihre Beantwortung geeignet ist. Allerdings ist das Kriterium der Angemessenheit insgesamt unscharf und bleibt somit weitgehend der Deutung jedes einzelnen Forschenden überlassen. Da es jedoch für den methodisch-methodologischen Diskurs unverzichtbar ist, ein konsensfähiges Begriffsverständnis zu haben, sollen im Folgenden diesbezügliche Überlegungen angestellt werden. Sie basieren auf den Recherchen von Stefer (2013), der die einschlägige forschungsmethodologische Literatur im Hinblick auf ihre Thematisierung der Angemessenheit von Methoden ausgewertet und dabei eine Reihe von Faktoren identifiziert hat, von denen die mir am wichtigsten erscheinenden im Folgenden grob skizziert werden sollen.

Zunächst einmal ist es unverzichtbar, dass der *Gegenstand* inklusive seiner Spezifika präzise benannt und definiert wird. Da die Auswahl von Methoden auch von der verfolgten *Fragestellung* und ihrer *Zielsetzung* abhängig ist, müssen auch diese möglichst genau und eindeutig formuliert werden: Sollen verallgemeinerbare Ergebnisse oder dichte Einzelfallbeschreibungen geliefert werden? Welche Methoden angewendet werden (können), hängt natürlich nicht zuletzt auch von den *Forschungsteilnehmer/innen* und ihren Eigenschaften, Kompetenzen und Bereitschaften ab sowie davon, welches Bild Forscher/innen von ihnen haben und welche Erwartungen damit verknüpft sind. Sind es Erwachsene, Jugendliche oder Kinder? Handelt es sich um Muttersprachler oder um Nicht-Muttersprachler der untersuchten Zielsprache? Über welchen Bildungshintergrund verfügen die Probandinnen und Probanden? Können also z.B. Befragungen bzw. Introspektionen durchgeführt werden, oder muss

sich die Forschung auf Beobachtungen und die Dokumentation und Analyse von Produkten beschränken? So kann man als empirische Fremdsprachenforscherin zwar ein ausgeklügeltes und aktuelle forschungsmethodologische Entwicklungen berücksichtigendes Design entwerfen, mit dem man z.B. interkulturelle Kompetenzen von Schülerinnen und Schülern in der Sekundarstufe, Fähigkeitsselbstkonzepte von Lehramtsstudierenden oder subjektive Lehrertheorien über das selbstgesteuerten Lernen im Fremdsprachenunterricht umfassend und vertiefend untersuchen kann. Wenn aber dabei die Protagonisten (also die Forschungsteilnehmer/innen) mit ihren kognitiven, affektiven oder sozialen Spezifika nicht in angemessener Weise berücksichtigt werden, ist das Scheitern solcher empirischer Studien vorprogrammiert. Hier kommt auch die von Appel und Rauin (in diesem Band) angesprochene Problematik der Invasivitätseffekte ins Spiel: Ein zu differenziertes Datenerhebungsdesign und die daraus resultierenden technischen Apparaturen können zu Verfälschungen, mindestens aber zu Verzerrungen und somit zu Beeinträchtigungen der Validität einer Studie führen. Nicht zu unterschätzen ist auch die mögliche zeitliche und je nach Thema ggf. auch psychische Belastung, die eine empirische Forschung für die Forschungsteilnehmer/innen bedeuten kann. Hier kommen also auch grundsätzliche ethische Fragen ins Spiel wie z.B.: Was kann/darf Forschungsteilnehmer/innen zugemutet werden (vgl. auch Viebrock 2009, Bach/Viebrock 2012)?

Welche Art von Forschung durchgeführt wird und welche empirischen Methoden angewendet werden, hängt ferner auch vom Interesse der *Adressaten* einer Untersuchung und ihrer Ergebnisse ab. Handelt es sich um ein wissenschaftliches Fachpublikum, das an inhaltlichen und methodischen Details interessiert ist und sich ggf. für die Replikation der vorgestellten Studie interessiert, oder handelt es sich z.B. um eine Zuhörerschaft, die einen allgemeinen Überblick z.B. über ein ihnen eher unvertrautes Thema wünscht, mit dem sie sich u.U. nicht weiter befassen werden, so dass ein allgemeines methodisches Instrumentarium ausreichend ist.

Ein weiterer Faktor zur Beurteilung der Angemessenheit ist die *Genese* der Methode. So stellen Flick, von Kardorff und Steinke (2000: 22) fest, dass sich für fast jedes Verfahren zurückverfolgen lässt, „für welchen besonderen Forschungsgegenstand es entwickelt wurde. Ausgangspunkt war in der Regel, dass die vorhandenen Methoden für diesen spezifischen Gegenstand nicht geeignet waren." Die Kenntnis der Historie einer Methode ermöglicht nicht nur ein besseres Verständnis ihrer Besonderheiten, sondern eben auch eine Einschätzung hinsichtlich der Frage eines möglichen Transfers auf einen Gegenstand, für den sie nicht originär entwickelt worden ist. Insgesamt ist festzustellen, dass in den

meisten qualitativen Forschungen hochgradig individuelle Designs für die bestmögliche Erfassung des zu untersuchenden Gegenstands entworfen werden. Dies bringt zwar den Vorteil der optimalen Passung mit sich, eine solche Vorgehensweise hat aber den Nachteil, dass Replikationsstudien nur sehr schwer oder ggf. gar nicht möglich sind. Dies ist eine in der forschungsmethodologischen Diskussion immer wieder formulierte Kritik an qualitativen Designs, insbesondere solchen, die triangulativ arbeiten.

Ein im vorliegenden Zusammenhang zentraler Faktor betrifft das *Verhältnis von Aufwand, Ressourcen und Nutzen*. Insbesondere Forschungsnovizen – aber nicht nur sie – unterschätzen häufig den Aufwand, der mit der Durchführung, der Aufbereitung und der Auswertung z.B. von qualitativen Interviews, Videomitschnitten fremdsprachenunterrichtlicher Interaktionen oder Erfassungen von L2-Erwerbsprozessen verbunden ist. Solche Fehleinschätzungen führen nicht selten zu den gefürchteten Datenfriedhöfen, d.h. große Mengen an erhobenen, z.T. sogar aufbereiteten Daten werden meist aus zeitlichen und ressourciellen Gründen nicht ausgewertet und so zu ‚toten' Daten. Der immense Aufwand, der insbesondere in mehrmethodisch angelegten qualitativen Studien von den Forschenden betrieben werden muss, wird auch in nahezu allen Beiträgen im vorliegenden Sammelband thematisiert. Im Falle von Forscherteams kann der mit der Triangulationsstrategie verbundene Aufwand zwar verteilt, aber nicht grundsätzlich verringert werden.

Aufgrund der Tatsache, dass sich die Verknüpfung qualitativer und quantitativer Daten immer mehr zur Normalität sozialwissenschaftlicher und damit auch fremdsprachendidaktischer Forschung entwickelt und sich gleichzeitig die technischen Möglichkeiten der Datenverarbeitung enorm verbessert haben, werden in letzter Zeit auch zunehmend Software-Programme zu computergestützten Analyse großer Datenmengen entwickelt und erfolgreich eingesetzt (vgl. die ausführliche Darstellung von Stefer, in diesem Band). Die Geschwindigkeit und Effizienz, mit der elektronisch erfasste Daten inzwischen verarbeitet werden können, ist beachtlich, aber auch sehr verführerisch, da grundsätzlich die Gefahr besteht, dass dann möglicherweise überflüssige Auswertungen nur deshalb vorgenommen werden, weil es möglich ist, nicht aber weil es sinnvoll ist. Die hier skizzierte Problematik ist nicht zu unterschätzen und sollte offener diskutiert werden, auch um gemeinsam Kriterien zu entwickeln, die als Entscheidungshilfen fungieren können, ob und inwiefern sich der mit einem mehrmethodischen Vorgehen verbundene Aufwand (wie z.B. Einarbeitung in verschiedene Theorien und Methoden, Einbeziehung verschiedener Kolleginnen und Kollegen, Konzeption und Durchführung longitudinaler Studien) tatsächlich lohnt, d.h. den erhofften Erkenntnisgewinn bringt.

Eng mit dem vorangegangenen Faktor verknüpft ist der folgende, nämlich die *Forschungskompetenzen* der jeweiligen Wissenschaftler. Wenngleich es ideal wäre, so kann sich selbstverständlich nicht jeder empirisch Forschende in allen für die Untersuchung eines Gegenstands relevanten bzw. geeigneten Methoden auskennen. Mögliche Konsequenzen könnten darin bestehen, dass man sich selbst weiterbildet und methodisch-methodologische Kompetenzen in bis dahin unbekannten Verfahren erwirbt oder dass man Experten – seien es Kolleginnen oder Kollegen, Experten aus einschlägigen Disziplinen oder universitäre Forschungsreferate – zu Rate zieht und um Unterstützung bittet. Außer bei Nachwuchswissenschaftlern, die sich ihr methodologisch-methodisches Rüstzeug in der Regel erst noch erarbeiten und dabei häufig außerordentliche und hochaktuelle Kompetenzen erwerben, werden in den meisten Fällen diejenigen Methoden ausgewählt und eingesetzt, hinsichtlich derer Forschende bereits über sichere Kenntnisse und Expertisen verfügen. Da wir es in der qualitativen Fremdsprachenforschung mit komplexen Gegenständen zu tun haben, stellt sich in Bezug auf das Kriterium der Angemessenheit nicht zuletzt auch die Frage nach der *Offenheit* der Methode: Ermöglichen sie a) die ganzheitliche Erfassung des Gegenstands und b) das Hinterfragen von Vorannahmen der Forschenden? Oder erfassen sie nur einen Ausschnitt des Gegenstands, so dass die Einbeziehung weiterer Methoden indiziert ist?

Resümierend lässt sich festhalten, dass alle genannten Faktoren hinsichtlich der Frage nach der Angemessenheit bzw. der Passung von Methoden für eine empirische Forschung zu berücksichtigen sind, wobei den Faktoren *Gegenstand* und *Fragestellung* höchste theoretische Priorität zukommt, während den Faktoren *Eigenschaften von Forschungsteilnehmerinnen und -teilnehmern*, *Verhältnis von Aufwand, Ressourcen und Nutzen* sowie *methodisch-methodologische Kompetenzen der Forschenden* höchste praktische Priorität beizumessen ist.

4. Fazit

Die Anwendung der Triangulationsstrategie ist Ausdruck eines gewachsenen Bewusstseins über die Relativität von Perspektiven, d.h. seitdem anerkannt wurde, dass nicht von einer objektiven Realität ausgegangen werden kann und es immer mehrere Perspektiven auf die Wirklichkeit und ihre Gegenstände gibt, erscheint es sinnvoll und wichtig, diese auch zu berücksichtigen, um eine halbwegs adäquate Rekonstruktion des jeweils fokussierten Wirklichkeitsausschnitts vorzunehmen.

Ein Ziel des vorliegenden Beitrags bestand darin zu zeigen, dass der Einsatz mehrerer Methoden innerhalb einer Forschung und die Anwendung der

Triangulationsstrategie nicht zwingend dasselbe sind. Im Rahmen einer Untersuchung mehrere Methoden, Verfahren etc. kumulativ einzusetzen, ohne sich über deren genaue Funktion und über ihre Wechselwirkungen Gedanken zu machen, kann zwar – vielleicht aber nur zufällig – zu relevanten Ergebnissen führen, eine wissenschaftliche Vorgehensweise im Sinne einer systematischen, begründeten und kontrollierten Anwendung forschungsmethodischer Verfahren ist damit jedoch nicht erreicht. Was zur Steigerung der Qualität empirischer Studien beiträgt, sind transparente, nachvollziehbare, strukturierte und fundierte Entscheidungen hinsichtlich der Auswahl und des Einsatzes von Methoden, und zwar einerseits sowohl bei der Datenerhebung, bei der Datenaufbereitung als auch bei der Datenauswertung und andererseits sowohl bei der Anwendung einer einzelnen Methode als auch beim Einsatz mehrerer Methoden. Grundsätzlich gilt jedoch, dass solange Grundprinzipien wie Angemessenheit, Systematizität und Nachvollziehbarkeit eingehalten werden, kein Verfahren wissenschaftlicher als ein anderes ist. Die methodologische Diskussion um eine systematisierte Kombination steht erst am Anfang: Sie muss weitergehen!

Falls Validierung überhaupt möglich ist, dann kann sie nur durch die Anwendung der Triangulationsstrategie erfolgen. Aus meiner Sicht besteht die Funktion der Triangulation in dem Versuch, verbesserte Erkenntnisse zu erzielen. Dass dieser Versuch zeitweise zu einer Modeerscheinung degeneriert ist – wie es z.B. Lamnek (2005) beklagt –, mag zwar stimmen, ist aber kein Grund, die Forschungsstrategie als Ganze in Zweifel zu ziehen. Allerdings ist zu beachten, dass unterschiedliche Gegenstände unterschiedliches Triangulationspotenzial haben, d.h. nicht alle Fragestellungen lassen alle Triangulationstypen zu.

Der kombinierte Einsatz von mehreren Methoden ist eine inzwischen so weit verbreitete und populäre Vorgehensweise in der empirischen Forschung, dass Tashakkori/Teddlie (2010: ix) von einem *third methodological movement* sprechen. Ob und wie sich diese Bewegung weiterentwickelt, bleibt abzuwarten. Wenn allerdings nicht in absehbarer Zeit einige der immer wieder aufgeworfenen kritischen Aspekte bearbeitet werden, steht zu befürchten, dass es keine konstruktive Weiterentwicklung geben wird. So bedarf u.a. die folgende Frage einer Klärung: Welche Methoden können/dürfen miteinander kombiniert werden, um einen möglichst großen Erkenntnisgewinn zu bieten?

Ein in jüngster Zeit immer wieder formuliertes Desiderat betrifft die Bildung von Forschungsverbünden oder zumindest die kollegiale bzw. kooperative Erstellung und Nutzung gemeinsamer Korpora oder Datenbanken. Hier geschieht leider nach wie vor viel zu wenig, was nicht zuletzt damit zusammenhängt, dass viele empirische Daten im Rahmen von Qualifikationsarbeiten erhoben und aufbereitet werden, die von den Autorinnen und Autoren dann als ‚ihre' Daten

betrachtet werden. Dies ist angesichts des enormen Aufwands, der damit verbunden sein kann, nachvollziehbar. Für die Weiterentwicklung empirischer Disziplinen wie der Fremdsprachenforschung wäre die ‚polyvalente' Nutzung von bereits vorhandenen Daten jedoch ein enormer Fortschritt, nicht zuletzt deshalb, weil er dazu beiträgt, Zeit und Kosten zu sparen. Es soll an dieser Stelle jedoch nicht verschwiegen werden, dass die Schaffung gemeinsamer Korpora z.B. aufgrund der nicht allen Bereichen gleichermaßen standardisierten Datenaufbereitung (vgl. z.B. das Problem unterschiedlicher Transkriptionskonventionen) kein trivialer Prozess ist, der meiner Meinung dennoch angegangen werden sollte, um langfristig die vorhandenen Ressourcen besser nutzen zu können.

Unterrichtsbeobachtungen sind aufwendig und komplex, ermöglichen jedoch Einsichten in interaktive Lehr- und Lernprozesse, die auf keine andere Weise zu gewinnen sind. Klieme und Rakoczy (2008: 225) fassen das Ziel der Unterrichtsforschung folgendermaßen zusammen: „Um wissenschaftliche Erkenntnisse und praktisch handhabbares Wissen zu erzeugen, bedarf es einer systematischen Beobachtung und Beschreibung der Interaktionsprozesse von Lehrern und Schülern sowie der Analyse ihres Zusammenhangs mit Schülermerkmalen (Lernvoraussetzungen, -strategien und -ergebnissen) und Lehrermerkmalen (z.B. allgemein-pädagogischem und fachdidaktischem Wissen), kurz: einer empirischen Unterrichtsforschung". Inzwischen gibt es zahlreiche technische Möglichkeiten, die eine solche faktorenkomplexe Forschung unterstützen können. Die im Zusammenhang z.B. mit der Videografie inzwischen mögliche Perspektivenvielfalt, d.h. der Blick auf dieselbe Situation aus verschiedenen Perspektiven, kann einen echten Erkenntnisgewinn ermöglichen – hier erscheint die Anwendung der Triangulationsstrategie also besonders fruchtbar. Das vorhandene Potenzial wird in den Beiträgen von Seifert sowie von Appel und Rauin (beide in diesem Band) überzeugend dargestellt. Zusätzlich zu den verschiedenen Perspektiven können auch zusätzliche Daten (Lautes Erinnern, Befragungen, Tests) erhoben werden, die zu einem besseren Verständnis der Videodaten beitragen und aus meiner Sicht eine hervorragende Möglichkeit darstellen, die beiden Funktionen der Triangulation – Validierung und Vertiefung – miteinander zu kombinieren. Die beispielsweise im Falle der Kombination von *online* oder *offline* durchgeführten Beobachtungen und anschließenden retrospektiven Befragungen (lautes Erinnern bzw. *stimulated recall*) ermöglichte Ermittlung der Innensicht von Forschungsteilnehmern bzw. die Rekonstruktion ihrer Wirklichkeitskonstruktion lassen Erkenntnisse zu, die mittels einer Einzelmethode nicht erzielbar wären.

In Bezug auf die in der Fremdsprachenforschung in jüngster Zeit vermehrt durchgeführten videobasierten Beobachtungsstudien teile ich daher das von Appel und Rauin (in diesem Band: 76) formulierte Fazit, dass sie „nicht nur

Möglichkeiten zur Triangulation eröffnen, sondern sie in Gestalt der Verknüpfung und Kombination von Daten, der Variation von Instrumenten und der Beteiligung mehrerer Forscher regelrecht erfordern, um Beobachtungsdaten in ihrem Aussagegehalt aufwerten zu können", wobei eine bewusste und reflektierte Auseinandersetzung mit den Verfahren unverzichtbar ist.

Literatur

Aguado, Karin (2009). Möglichkeiten und Grenzen mehrmethodischer empirischer Fremdsprachenlehr- und -lernforschung. In: Baumann, Beate / Nied, Martina / Hoffmann, Sabine (Hrsg.). *Qualitative Forschung in Deutsch als Fremdsprache*. Frankfurt/Main: Lang, 13-22.

Aguado, Karin (2014). Triangulation. In: Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hrsg.). *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung.* Paderborn: Schöningh, 47-56.

Aguado, Karin / Riemer, Claudia (2001). Triangulation: Chancen und Grenzen mehrmethodischer empirischer Forschung In: Aguado, Karin; Riemer, Claudia (Hrsg.). *Wege und Ziele. Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen). Festschrift für Gert Henrici.* Baltmannsweiler: Schneider-Verlag Hohengehren, 245-257.

Appel, Johannes / Rauin, Udo (2014). Methoden videogestützter Beobachtungsverfahren in der Lehr-Lern-Forschung. In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 59-79.

Arbeitsgruppe Bielefelder Soziologen (Hrsg.) (1976). *Kommunikative Sozialforschung. Alltagswissen und Alltagshandeln, Gemeindemachtforschung, Polizei, politische Erwachsenenbildung*. München: Fink.

Bach, Gerhard / Viebrock, Britta (2012). Was ist erlaubt? Forschungsethik in der Fremdsprachenforschung. In: Doff, Sabine (Hrsg.). *Fremdsprachenunterricht empirisch erforschen: Grundlagen, Methoden, Anwendung*. Tübingen: Narr, 19-35.

Birkner, Hans Albert / Hackfort, Dieter (2006). Triangulation als Grundlage diagnostischen Urteilens. *Zeitschrift für Sportpsychologie* 13/2, 75-80.

Blaikie, Norman W. (1991). A Critique of the Use of Triangulation in Social Research. *Quality and Quantity* 25/2, 115-136.

Brüsemeister, Thomas (2008). *Qualitative Forschung. Ein Überblick*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

Campbell, David T. / Fiske, Donald W. (1959). Convergent and Discriminant Validation by the Multitrait-multimethod Matrix. *Psychological Bulletin* 56, 81-105.

Dallinger, Sara / Jonkmann, Kathrin (2014). Competencies and Motivation in the Bilingual History Classroom (COMBIH) – Eine Längsschnittstudie zu deutsch-englischem Geschichtsunterricht an Gymnasien. In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 181-201.

Denzin, Norman K. (1970). The Research Act. A Theoretical Introduction to Sociological Methods. Chicago: Aldine.

Denzin, Norman K. / Lincoln, Yvonna S. (1994). Introduction: Entering the Field of Qualitative Research. In: Denzin, Norman K. / Lincoln, Yvonna S. (Hrsg.). *Handbook of Qualitative Research.* Thousand Oaks u. a.: Sage, 1-17.

Elsner, Daniela / Viebrock, Britta (2014). Einleitung: Triangulation in der Fremdsprachenforschung. In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 7-16.

Fielding, Nigel G. / Fielding, Jane L. (1986). *Linking Data.* Beverly Hills: Sage.

Flick, Uwe (2004). *Triangulation. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (2000). Was ist qualitative Forschung. Einführung und Überblick. In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch.* 3. Aufl. Reinbek bei Hamburg: Rowohlt, 13-29.

Geertz, Clifford (1983). *Dichte Beschreibung.* Beiträge zum Verstehen kultureller Systeme. Frankfurt/Main: Suhrkamp.

Grotjahn, Rüdiger (2006). Zur Methodologie der Fremdsprachenerwerbsforschung. In: Scherfer, Peter / Wolff, Dieter (Hrsg.). *Vom Lehren und Lernen fremder Sprachen: Eine vorläufige Bestandsaufnahme*. Frankfurt/Main: Lang, 247-270.

Ivankova, Nataliya. V. / Creswell John W. (2009). Mixed Methods. In: Heigham, J. / Croker, R. A. (Hrsg). *Qualitative Research in Applied Linguistics: A Practical Introduction*. Basingstoke: Palgrave Macmillan, 135-161.

Kelle, Udo / Erzberger, Christian (2000). Qualitative und quantitative Methoden: kein Gegensatz. In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (Hrsg.). *Qualitative Forschung. Ein Handbuch.* 3. Aufl. Reinbek bei Hamburg: Rowohlt, 299-309.

Klieme, Eckhard / Rakoczy, Katrin (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik* 54/2, 222-237.

Kuckartz, Udo (2014). *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lamnek, Siegfried (2005). *Qualitative Sozialforschung. Lehrbuch.* 4. Aufl. Weinheim/Basel: Beltz.

Mayring, Philipp (2001). Kombination und Integration qualitativer und quantitativer Analyse. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* 2, Art. 6. Online: http://nbn-resolving.de/urn:nbn:de:0114-fqs010162 [10. 3.2014].

Mayring, Philipp (2002). *Einführung in die qualitative Sozialforschung.* Weinheim/Basel: Beltz.

Schütz, Alfred (1974). *Der sinnhafte Aufbau der sozialen Welt. Eine Einleitung in die verstehende Soziologie.* Frankfurt/Main: Suhrkamp

Seifert, Heidi (2014). Videografie als Instrument zur Erforschung von Interaktionsprozessen im Elementarbereich. In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 81-100.

Settinieri, Julia (2014). Forschst Du noch, oder triangulierst Du schon? In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 17-35.

Silverman, David (1985). *Qualitative Methodology and Sociology*. Aldershot: Gower.

Stefer, Claus (2013). *Die Gegenstandsangemessenheit empirischer Datenerhebungsmethoden im Kontext von Lehrevaluationen an Hochschulen*. Dissertation am Fachbereich Erziehungswissenschaften, Philipps-Universität Marburg/Lahn.

Stefer, Claus (2014). Ansätze zur Triangulation mithilfe qualitativer Forschungssoftware (MAXQDA 11). In: Elsner, Daniela / Viebrock, Britta (Hrsg.). *Triangulation in der Fremdsprachenforschung*. Frankfurt/Main: Lang, 101-133.

Tashakkori, Abbas / Teddlie, Charles (2003). Major Issues and Controversies in the Use of Mixed Methods in Social and Behavioral Research. In: Tashakkori, Abbas / Teddlie, Charles (Hrsg.). *Handbook of Mixed Methods in Social and Behavioral Research.* 2. Aufl. Thousand Oakes, CA: Sage, ix-xv.

Viebrock, Britta (2009). Fremdsprachenforschung und *Research Ethics*. In: Lütge, Christiane / Kollenrott, Anne Ingrid / Ziegenmeyer, Birgit / Fellmann, Gabriela (Hrsg.).*Empirische Fremdsprachenforschung – Konzepte und Perspektiven*. Frankfurt/Main: Lang, 39-59.

Anschriften der Autorinnen und Autoren

Prof. Dr. Karin Aguado, karin.aguado@uni-kassel.de
Universität Kassel, Institut für Germanistik, Kurt-Wolters-Straße 5, 34125 Kassel

Johannes Appel, appel@em.uni-frankfurt.de
Goethe-Universität Frankfurt, Akademie für Bildungsforschung und Lehrerbildung (ABL), Senckenberganlage 31, 60054 Frankfurt/Main

Laura Armbrust, Laura.Armbrust@gmx.de

Sonja Brunsmeier, brunsmeier@ph-freiburg.de
PH Freiburg, Institut für Anglistik, Kunzenweg 21, 79117 Freiburg

Sara Dallinger, dallinger@ph-ludwigsburg.de
Pädagogische Hochschule Ludwigsburg, Institut für Sprachen, Abteilung Englisch, Reuteallee 46, 71634 Ludwigsburg

Prof. Dr. Daniela Elsner, elsner@em.uni-frankfurt.de
Goethe-Universität Frankfurt/Main, Institut für England- und Amerikastudien, Grüneburgplatz 1, 60629 Frankfurt/Main

Prof. Dr. Kathrin Jonkmann, Kathrin.Jonkmann@FernUni-Hagen.de
FernUniversität in Hagen, Institut für Psychologie, LG Bildungspsychologie, Universitätsstr. 33, 58097 Hagen

Dr. Astrid Jurecka, jurecka@em.uni-frankfurt.de
Goethe-Universität Frankfurt, Institut für Pädagogik der Elementar- und Primarstufe, Grüneburgplatz 1, 60629 Frankfurt/Main

Prof. Dr. Udo Rauin, rauin@em.uni-frankfurt.de
Goethe-Universität Frankfurt, Akademie für Bildungsforschung und Lehrerbildung (ABL), Senckenberganlage 31, 60054 Frankfurt/Main

Heidi Seifert, hseifert@spz.tu-darmstadt.de
Technische Universität Darmstadt, Institut für Sprach- und Literaturwissenschaft, Hochschulstraße 1, 64289 Darmstadt

Prof. Dr. Julia Settinieri, jsetti@mail.uni-paderborn.de
Universität Paderborn, Institut für Germanistik und Vergleichende Literaturwissenschaft, Warburgerstraße 100, 33100 Paderborn

Dr. Claus Stefer, stefer@uni-marburg.de
Philipps-Universität Marburg, Institut für Erziehungswissenschaft, Bunsenstraße 3, 35037 Marburg

Prof. Dr. Britta Viebrock, viebrock@em.uni-frankfurt.de
Goethe-Universität Frankfurt/Main, Institut für England- und Amerikastudien, Grüneburgplatz 1, 60629 Frankfurt/Main